

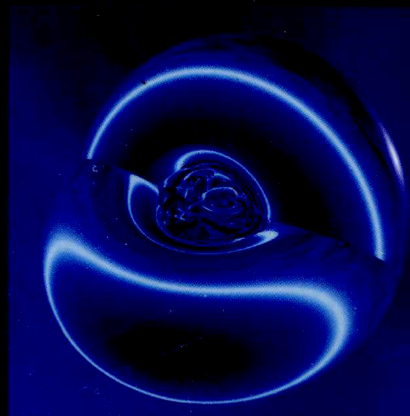
国际信息工程先进技术译丛



认知无线电网络

COGNITIVE RADIO NETWORKS

（中国） Kwang-Cheng Chen
（丹麦） Ramjee Prasad 著
许方敏 李虎生 译





机械工业出版社
CHINA MACHINE PRESS



国际信息工程先进技术译丛

认知无线电网络

（中国）Kwang-Cheng Chen
（丹麦）Ramjee Prasad 著
许方敏 李虎生 译



机 械 工 业 出 版 社

本书系统地介绍了认知无线电网络及其关键技术，从无线通信基础技术（包括正交频分复用和多输入多输出技术）和无线网络基础（多址接入、路由等）引入，按照物理层、媒体接入控制层和网络层的顺序，分别探讨了频谱感知技术、媒体接入控制层协议、路由算法、安全性和频谱管理方法等。相比其他认知无线电的书籍，本书重点讨论了网络层相关技术。

本书是目前唯一对认知无线电网络提供指导和参考的书籍，本书提供的背景知识适合于初学者作为掌握无线通信以及无线网络的参考书，而后续章节也适合于想深入了解和研究认知无线电网络的研究者。

图字：01-2009-6679 号。

图书在版编目（CIP）数据

认知无线电网络/（中国）陈光桢，（丹麦）普拉萨德（Prasad，R.）著；许方敏，李虎生译. —北京：机械工业出版社，2010.12

（国际信息工程先进技术译丛）

COGNITIVE RADIO NETWORKS

ISBN 978-7-111-32040-1

Ⅰ.①认… Ⅱ.①陈…②普…③许…④李… Ⅲ.①无线电通信－通信网 Ⅳ.①TN92

中国版本图书馆 CIP 数据核字（2010）第 190088 号

机械工业出版社（北京市百万庄大街 22 号　邮政编码 100037）

策划编辑：张俊红　责任编辑：闾洪庆　版式设计：霍永明

责任校对：张晓蓉　封面设计：马精明　责任印制：李　妍

北京外文印刷厂印刷

2011 年 1 月第 1 版第 1 次印刷

169mm×239mm·23 印张·448 千字

0001—3000 册

标准书号：ISBN 978-7-111-32040-1

定价：88.00 元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换

电话服务

社服务中心：（010）88361066

销 售 一 部：（010）68326294

销 售 二 部：（010）88379649

读者服务部：（010）68993821

网络服务

门户网：http://www.cmpbook.com

教材网：http://www.cmpedu.com

封面无防伪标均为盗版

译 者 序

认知无线电及认知网络是近几年无线通信领域出现的重要新课题之一，也是当今无线通信领域研究的最新热点。

在第四代移动通信标准化即将完成之时，人们都在思考下一代无线通信应该包含哪些新技术、新突破。尤其是近年来，信道编码、多址接入、多天线传输等传统方法进展不大，很难满足越来越多的无线设备和业务需求，有限的频谱资源正成为无线通信发展的最大瓶颈之一，认知无线电的出现正好解决了固定频谱分配带来的频谱利用效率低下的问题。

由于认知无线电技术是近年来出现的新技术，因此目前全面系统论述认知无线电的书籍寥寥可数，而介绍认知无线电网络的书籍就更少了。本书是一本全面反映认知无线电网络最新技术的佳作。本书的两位作者都是从事无线通信理论和技术研究多年的教授，因此本书没有局限在认知无线电技术本身，而是突出认知无线电和无线网络的背景介绍，强调对读者的启发和引导，循序渐进，由浅入深，在内容上反映了当前的最新技术和研究成果。

作为一本介绍认知无线电技术和认知无线电网络的书籍，本书不仅可以面向初学者作为入门的参考书，也适合于想深入了解和研究认知无线电网络的研究人员。

本书第1~5章、第11章由许方敏翻译，感谢三星通信研究院的李小强、郑旭峰、孙程君、何宏、雷海鹏、周雷、王海、惠俊红、仲川、李迎阳等同事，他们也参与了资料整理和部分翻译工作；第6~10章由李虎生翻译，感谢毛如坤、杨德鹏、张正浩、郑坤等同学，他们为本书翻译做出了贡献，参加了部分文字编辑工作。

此外，对所有关心和支持本书翻译工作的同事、朋友致以诚挚的感谢！翻译的过程对于每位译者也是一次学习和思考的历程。最后，限于译者自身的水平及经验，错漏和不足在所难免，恳请读者批评指正。

译 者

原书序

无线通信和网络在过去数十年内得到了迅速增长，每年有数十亿新无线设备投入使用。我们预计，在未来十年无线设备数量将出现指数式增长，最终将导致适合用于无线通信的频谱资源稀缺。与传统提高物理层传输频谱效率的方法不同，Joe Mitola Ⅲ博士基于软件无线电的概念提出认知无线电技术，认知无线电通过利用频谱“空洞”或“白空间”来提高频谱利用率。美国联邦通信委员会（FCC）很快就认识到认知无线电的潜在应用前景，并同意应用认知无线电技术。目前在IEEE论文库里已经有超过1000篇认知无线电技术方面的研究论文，这正说明了认知无线电技术的重要性。但是，研究者也逐渐发现认知无线电技术并不能保证无线网络传输数据在链路层的频谱效率，只有把认知无线电设备通过中继协作等与主要/合法的无线电系统共存使用，才能进一步提高频谱利用率。因此，考虑到这个技术发展方向，我们编写了这本关于认知无线电网络的书籍，介绍从认知无线电到认知无线电网络最新的研究进展和成果。

在准备撰写这本书的期间，作者要感谢很多国际研究者和我们的学生的鼓励、讨论和技术支持，包括Mohsen、Guizani、Fleming Bjerge Frederiksen、Neeli Prasad、Ying-Chang Liang、Sumei Sun、Songyoung Lee、Albena Mihovska、Feng-Seng Chu、Chi-Cheng Tseng、Shimi Cheng、Lin-Hung Kung、Chung-Kai Yu、Shao-Yu Lien、Sheng-Yuan Tu、Bilge Kartal Cetin、Yu-Cheng Peng、Jin Wang、Peng-Yu Chen、Chu-Shiang Huang、Ching-Kai Liang、Hong-Bin Chang、Po-Yao Huang、Wei-Hong Liu、I-Han Chiang、Michael Eckl、Yo-Yu Lin、Weng Chon Ao、Dua Idris，以及认知无线电之父Joe Mitola Ⅲ。我们还要感谢Inga、Susanne和Keiling在各方面提供的帮助。

第一作者（K. C. Chen）还要特别感谢Irving T. Ho基金会，该基金会向台湾大学提供资助，从而让作者能专注于本书的写作。Irving T. Ho博士是台湾新竹科技园的创始人。此外，作者还要感谢丹麦国家科学理事会和奥尔堡大学电信研究中心提供机会让KC和Ramjee在丹麦一起工作。最后，但并非是最不重要的，KC要感谢他的妻子Chirstine与他的孩子Chloé和Danny的支持，尤其是他在2008年夏天为完成该书稿而没有时间在家里陪伴他们。

Kwang-Cheng Chen（中国）
Ramjee Prasad（丹麦）

目　　录

第1章 无线通信

传统的无线通信网络使用电路交换，例如使用频分多址（FDMA）的第一代蜂窝高级移动电话系统（AMPS）和采用时分多址（TDMA）的第二代蜂窝全球移动通信系统（GSM），或IS-95码分多址（CDMA）系统。分组交换对于因特网的成功起着非常重要的作用，也使得分组交换几乎被所有的无线宽带通信技术所采用。从第三代移动通信系统开始，分组交换正日益成为一个被普遍采用的技术。

国际标准化组织（ISO）已经定义了大量的计算机网络标准，包括计算机网络的开放系统互连（OSI）七层基本架构。这种OSI七层协议对于优化网络效率也许不是最理想的方案，但是对于大规模的网络，使用这种分层结构会带来很多好处。工程人员可以独立开发很多网络的软件和硬件部分，甚至是即插即用式网络，只要定义好网络各层之间的接口和标准，就可以随意替换部分网络硬件和软件。考虑到分组交换网络“统计复用”的本质，OSI分层结构可以推动计算机网络和本书探讨的无线宽带通信网络的发展。

图1-1描述了OSI七层结构和在网络扩展以及与网络其他部分互连中的应

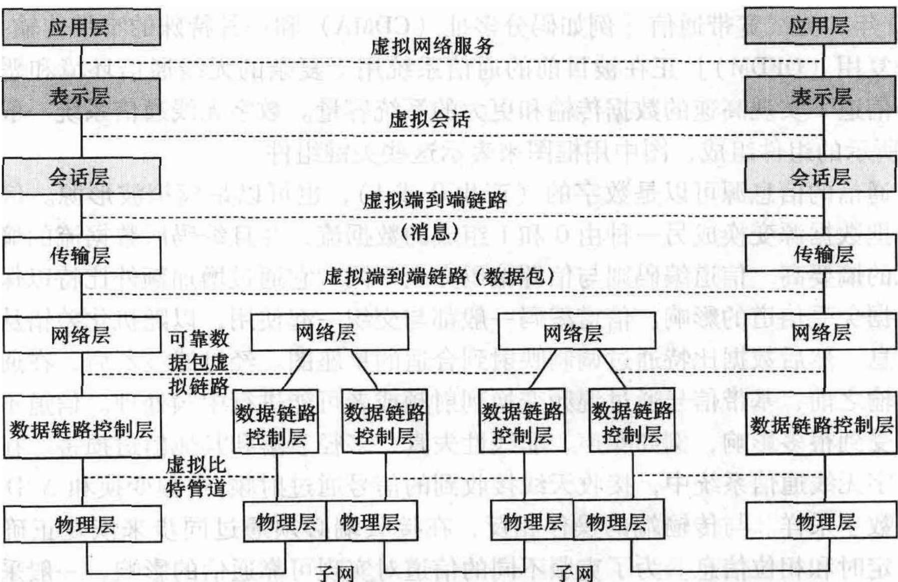


图1-1 OSI七层网络架构

用。OSI 中的上面四层在网络运营中主要是逻辑的概念而不是物理的概念，在OSI 最下面的两层（物理层和数据链路层）中传送、接收和协调实际的信令。无线网络的物理层通过无线媒介发射和准确接收数据比特，而媒体接入控制（MAC）层使用由数据比特组成的媒介来协调分组数据包的传输。我们在本书中讨论的无线通信，一般都是指无线网络的物理层和 MAC 层，而在其他书籍中无线通信可能包含更大的范围。在本章中，我们主要介绍无线通信系统的物理层传输和狭义的无线通信中的一些关键技术，也就是正交频分复用（OFDM）和多输入多输出（MIMO）处理。

1.1 无线通信系统

为了支持目前无线移动通信网络中的多媒体业务，数字通信系统工程已经被广泛用于物理层传输。为了能让后续章节更加易于理解，我们先简单介绍数字通信的基本原理，并且假设读者已经初步地了解一些通信系统和信令的知识。有兴趣的读者可以通过本章找到更深入的参考内容。

在模拟幅度调制（AM）和频率调制（FM）之后，数字通信系统已经在过去半个世纪中被广泛研究。由于系统链路性能更好，数字通信比模拟通信有着很多优势，此外数字通信比模拟通信传输更加可靠。在过去，假设电话线调制器已经达到理论极限，大多数研究都集中在传统的窄带数据传输。无线数字通信是随着许多新兴的应用（例如卫星通信和模拟蜂窝通信）兴起而得到发展的。在过去 20 年，无线宽带通信［例如码分多址（CDMA）和一种特殊的窄带传输-正交频分复用（OFDM）］正在被目前的通信系统用于复杂的无线通信环境和恶劣的衰落信道下实现高速的数据传输和更大的系统容量。数字无线通信系统一般由图 1-2 所示的组件组成，图中用框图来表示这些关键组件。

通信的信息源可以是数字的（产生 0 或 1），也可以是模拟波形源。信源编码器把数据源变换成另一种由 0 和 1 组成的数据流，并且编码后数据流的熵比数据源的熵要高。信道编码则与信源编码完全不同，它通过增加额外比特以保护原始数据免受信道的影响。信道编码一般都与交织一起使用，以随机化差错从而保护信息。然后数据比特通过调制映射到合适的星座图。经过滤波器后，在通过天线传输之前，基带信号通过混频变换到射频或者可能进行中频处理。信道不可避免会受到很多影响，例如噪声、非线性失真、多径衰落和其他信道损害。在现代的数字无线通信系统中，接收天线接收到的信号通过射频/中频变换和 A/D 转换变成数字采样。与传输端的操作相反，在接收端必须通过同步来恢复正确的频率、定时和相位信息。为了克服不同的信道对实现可靠通信的影响，一般采用信道均衡来弥补信道的影响。为了进一步保证可靠的通信和导频信号，在很多现代

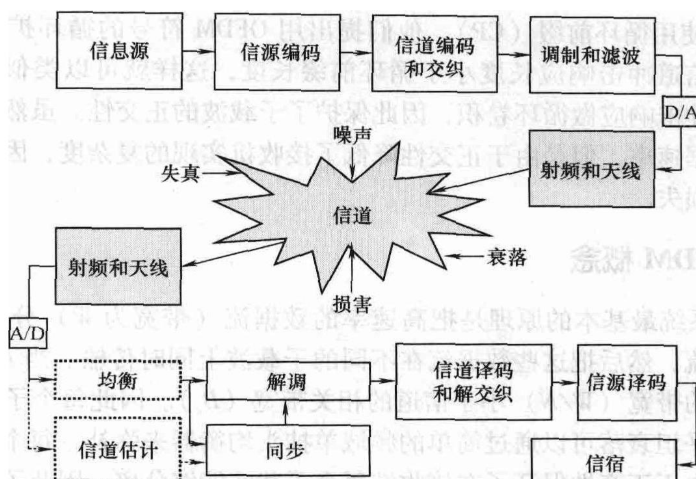


图 1-2　一个典型的数字无线通信系统框图

通信系统里使用信道估计来增强处理接收信号的能力。

总而言之，无线数字通信系统的物理层是为了对抗噪声和信道损害（信道、衰落、移动等造成的非线性失真）造成的符号间干扰（ISI）。目前的数字通信系统都是在硬件（例如集成电路）中或者运行在数字信号处理器或微处理器上的软件中实现这些功能的。

在本章 1.2 节中，我们将简单介绍 OFDM 和相应的多址接入方式——正交频分复用接入（OFDMA）。

1.2　正交频分复用

在 1960 年，Chang 在参考文献［1］中提出了在线性带限信道中同时传输信息，并且避免了信道间干扰和符号间干扰的原理。Saltzberg 在参考文献［2］中分析了上述系统的性能，得出的结论是有效的并行传输系统更需要考虑如何减少相邻信道的干扰，而不是让单个信道信号质量最好，因为相邻信道的干扰比每个信道内的干扰影响更大。这个结论在后来的基带数字信号处理中得到了验证。

OFDM 技术的主要贡献是 Weinstein 和 Ebert 在参考文献［3］中提出使用离散傅里叶变换（DFT）来进行基带调制和解调。使用 DFT 可以极大地提高调制和解调处理的效率，使用保护间隔和升余弦滤波能很大程度上解决符号间干扰的问题。虽然他们提出的系统并不能保证在时间扩展信道中子载波之间的正交性，但是它仍然对于 OFDM 系统的研究做出了很大的贡献。

为了解决弥散（衰落）信道中保持正交性的问题，Peled 和 Ruiz 在参考文献

［4］中提出使用循环前缀（CP），他们提出用 OFDM 符号的循环扩展填充保护间隔，只要信道冲击响应长度小于循环前缀长度，这样就可以类似于把 OFDM 符号与信道冲击响应做循环卷积，因此保护了子载波的正交性。虽然增加循环前缀会降低数据速率，但是由于正交性降低了接收机实现的复杂度，因此可以弥补数据速率的损失。

1.2.1 OFDM 概念

OFDM 系统最基本的原理是把高速率的数据流（带宽为 W）分解为 N 个低速率的数据流，然后把这些数据流在不同的子载波上同时传输。当 N 足够大时，每个子载波的带宽（W/N）小于信道的相关带宽（B_c）。因此每个子载波经过平坦衰落，而平坦衰落可以通过简单的频域单抽头均衡器来弥补。每个子载波之间相互正交，由于正交性保证了在接收端每个子载波能够分离，因此子载波可以在频域上相互重叠。与载波之间不能重叠的 FDMA 系统相比，OFDM 系统具有更高的频谱效率。

OFDM 系统的频谱效率如图 1-3 所示，图中给出了传统的非重叠多载波技术（例如 FDMA）和重叠多载波调制技术［例如离散多音频（DMT）、OFDM 等］的区别。如图 1-3 中所示（该图只是为了说明基本原理，实际的多载波技术如图 1-5 所示），使用重叠多载波调制技术可以实现更高的频谱利用率。但是，实现重叠多载波调制需要减少子载波之间的干扰，也就是应在子载波之间保证正交性。

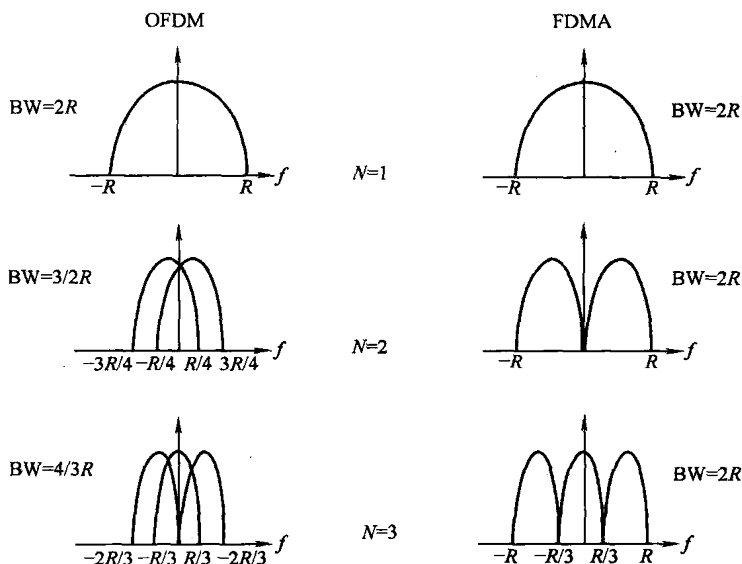


图 1-3　正交多载波和传统多载波

在基于 OFDM 的系统中，正交性要求子载波频率之间满足精确的数学关系。在一般的频分多址系统中，不同的载波在频域上分隔开来，信号可以使用传统的滤波器和解调器接收。在接收端，需要在载波间设置频域保护带，这样就导致频谱效率的浪费。然而在 OFDM 系统中，相邻子载波可以重叠部分边带，在接收端可以避免相邻子载波的干扰，并正确地接收信号。OFDM 接收端可以构建成解调器组，把每个子载波变换到直流，在一个符号周期内积分，从而恢复出原始数据。如果所有的子载波下变频后在时域上一个符号周期 T 内有整数个频率周期，那么积分后将没有载波间干扰。如果载波间间隔是 $1/T$ 的整数倍，这些子载波可以看成是线性独立（也就是正交），在后面的章节中证明了在基于 OFDM 的系统中其子载波之间的正交性。

图 1-4 给出了独立数据子载波的频谱，图 1-5 给出了一个 OFDM 符号的频谱。子载波间隔等于每个子载波的传输带宽的 OFDM 信号功率谱，如图 1-4 所示，图 1-5 的结果表明每个子载波的中心频率处没有相邻信道的干扰。因此，如果接收机在每个子载波的中心频率处进行相关，就能准确地恢复出传输的数据。此外，通过基于 DFT 的多载波技术，频分复用可以通过基带处理实现，而不是通过更昂贵更复杂的带通处理来实现。

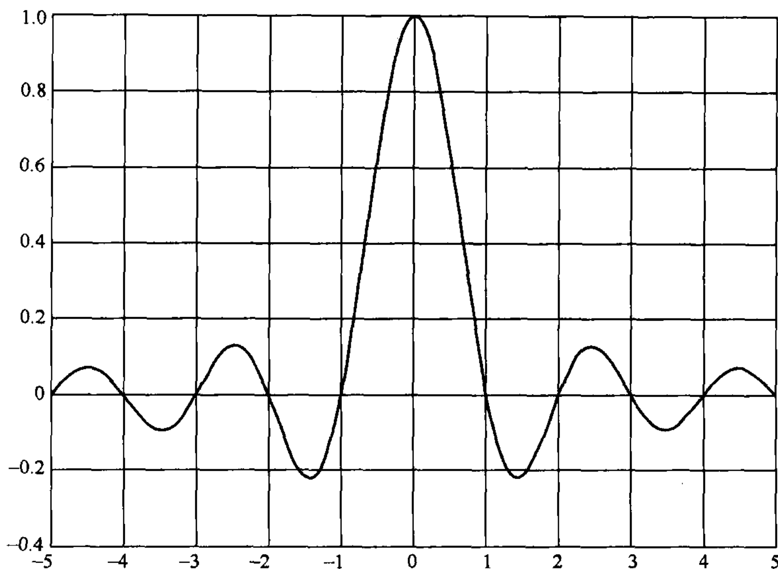


图 1-4 OFDM 独立子载波的频谱

在时间弥散信道中，是通过增加 CP 来保持子载波的正交性的。CP 是一个 OFDM 符号的最后一部分，一般把 CP 添加在传输的 OFDM 符号的起始部分，CP 可以用于对抗载波间干扰的影响。图 1-6a、b 分别给出了简化的 OFDM 系统发射机和接收机框图。

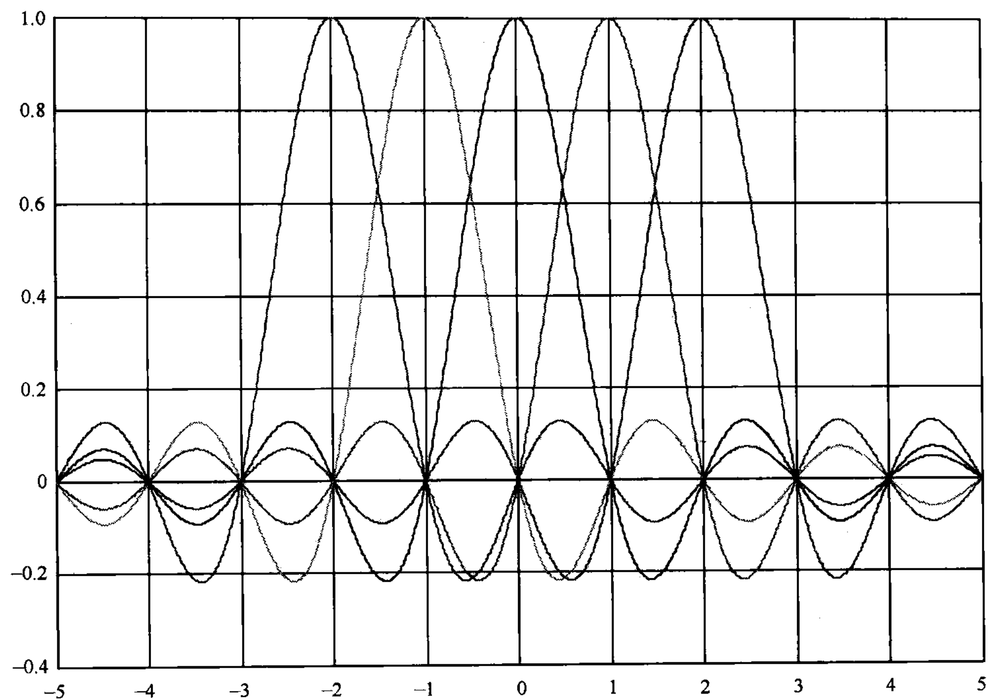
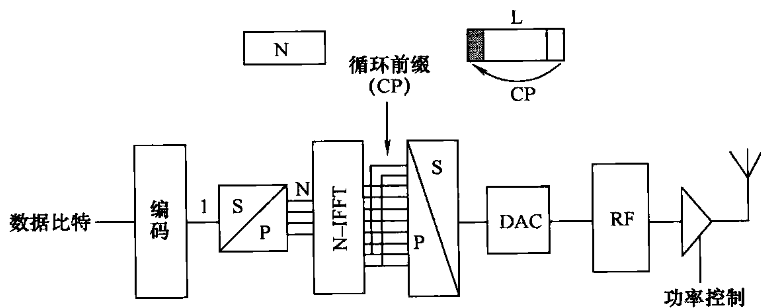
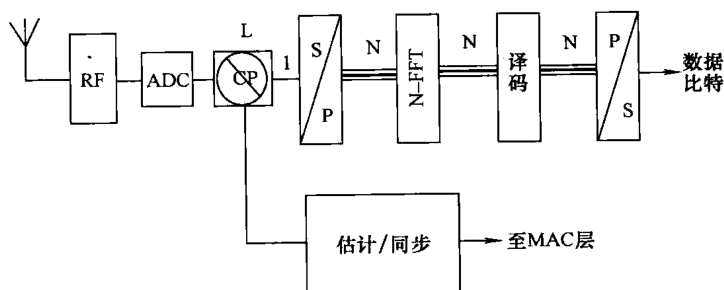


图 1-5　OFDM 符号的频谱



a)



b)

图 1-6　a）发射机框图和 b）接收机框图

1.2.2 OFDM系统的数学模型

如图1-7所示，基于OFDM的通信系统使用正交的子载波来同时传输多个数据码。出于简化的目的，图中没有给出用于抑制符号间干扰的保护间隔。数据符号 $d_{n,k}$ 首先变成一组大小为 N 的数据块，然后用复正则（本书中采用指数函数）波形 $\{\phi_k(t)\}_{k=0}^{N}$ 对数据进行调制［见式（1-1）］。调制后的数据在接收端以数据流的形式同时传输。图1-7所示的调制器可以使用快速傅里叶逆变换（IFFT）模块来实现，用式（1-1）描述如下：

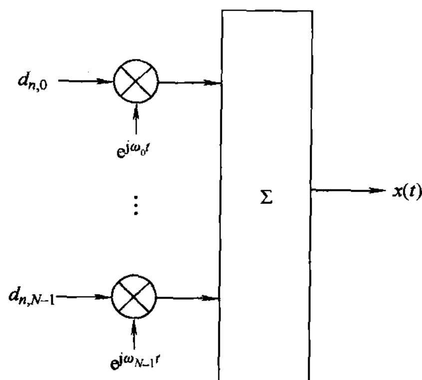


图1-7 OFDM调制器

$$x(t)=\sum_{n=-\infty}^{\infty}\left[\sum_{k=0}^{N-1} d_{n,k}\phi_k(t-nT_d)\right] \tag{1-1}$$

式中

$$\phi_k(t)=\begin{cases}e^{j2\pi f_k t}, & t\in[0, T_d]\\ 0, & 其他\end{cases}$$

并且

$$f_k=f_0+\frac{k}{T_d},\ k=0, 1, \cdots, N-1$$

我们使用了下列符号：

$d_{n,k}$：在第 n 个时间间隔使用第 k 个子载波传输的符号；

T_d：符号持续时间；

N：OFDM子载波数量；

f_k：第 k 个子载波频率，其中 f_0 是最低的子载波频率。

OFDM解调器简化的模块图如图1-8所示。解调的过程是根据子载波 $\{\phi_k(\tau)\}$

的正交性，即

$$\int_{\Re} \phi_k(t)\phi_l^*(t)\,\mathrm{d}t = T_d\delta(k-l) = \begin{cases} T_d, & k=l \\ 0, & \text{其他} \end{cases}$$

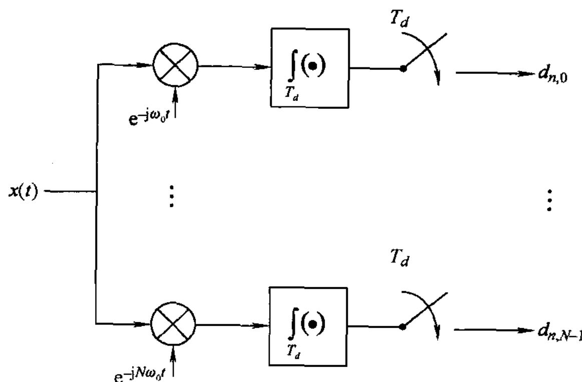


图 1-8 OFDM 解调器

因此，解调器可以利用子载波的正交性来数字实现，也就是利用快速傅里叶逆变换（IFFT）/快速傅里叶变换（FFT）实现 OFDM 信号的调制和解调：

$$d_{n,k} = \frac{1}{T_d}\int_{nT_d}^{(n+1)T_d} x(t) * \phi_k^*(t)\,\mathrm{d}t \tag{1-2}$$

式（1-2）可以使用图 1-8 中的快速傅里叶变换模块来实现。

上面的 OFDM 模型还可以用二维时域-频域网格图来描述，OFDM 的正交性可以用于弥补信号衰落等影响。图 1-7 中的 OFDM 调制器，可以找到一个表示 OFDM 传输信号模型的式（1-3）。此外，该特征还可以用于传输信号的脉冲成形以抑制符号间干扰和多径延迟扩展，详细解释如图 1-8 所示。

$$x(t) = \sum_{k,l} d_k \phi_{k,l}(t) \tag{1-3}$$

式中，$\phi_{k,l}(t)$ 表示基函数 $\phi(t)$ 分别在二维时域-频域网格图的时域和频域上延迟了 $l\tau_0$ 和 kv_0，如图 1-9 所示。可以看出，$\phi_{k,l}(t)$ 与基函数的关系可以用式（1-4）来表示

$$\phi_{k,l}(t) = \phi(t - l\tau_0)\mathrm{e}^{\mathrm{j}2\pi k v_0 t} \tag{1-4}$$

一般选择幅度为 $1/\sqrt{\tau_0}$、持续时间为 τ_0、频率间隔为 $v_0 = 1/\tau_0$ 的矩形脉冲为基函数 $\phi(t)$。网格图中每个传输的信号在传输过程中经历着同样的平坦衰落，因此简化了信道

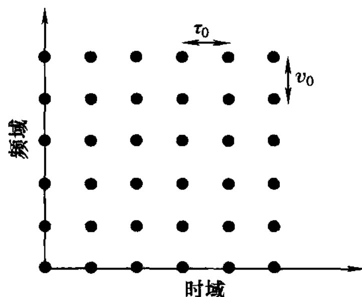


图 1-9 二维时频域网格图

估计和均衡过程。信道衰减是通过把接收信号与网格图上的先验符号做相关来估计的。这种技术通常用在基于 OFDM 的通信系统中来提供导频辅助信道估计。

1.2.3 OFDM 设计上的问题

基于 ODFM 的通信系统在频谱效率上有着与生俱来的优势，但是 OFDM 系统容易受到无线环境的影响。为了利用 OFDM 系统频谱效率和简单的发射机、接收机设计，这些信道的影响必须通过设计来处理，以获得潜在的增益。在通信系统中，接收机需要与发射机在频率、相位和时间（或帧/时隙/数据包边界）同步，从而正确地恢复出发射信号。在移动环境下，由于周围的信道变化太快，同步并不是一个特别简单的工作。例如，当移动设备开机时，移动设备并不知道周围的环境信息，它必须通过一些步骤（根据预定的协议/标准）来建立与基站/接入点之间的连接。在通信术语中，这个过程称为同步和获取。同步和获取的任务十分复杂，同时信道的影响使得这个任务更加困难。在以下章节中将详细探讨如何减少信道的影响。

1.2.3.1 频率偏移

OFDM 系统的频率偏移是由两个方面的因素造成的：发射机和接收机的采样时钟不匹配，发射站和接收站的参考频率不一致。下面将分析这两种频率偏移的原理和对系统性能的影响。

接收信号的采样时刻是由接收机中的模数采样时钟决定的，采样时钟有时不能完全与发射端的采样时钟匹配，从而造成接收采样时刻相对于发射采样时刻变慢。很多作者分析了采样时钟偏移对系统性能的影响。采样时钟偏差有两种表现形式：首先，采样时刻变慢会造成子载波旋转，进而因为载波间干扰（ICI）影响信噪比；此外，由于相邻子载波的能量扩展而导致子载波之间正交性变差。我们定义归一化采样误差为

$$t_\Delta = \frac{T' - T}{T}$$

式中，T 和 T'分别是发射和接收采样周期。在 DFT 后，接收机子载波 $R_{l,k}$上总的影响为

$$R_{l,k} = \mathrm{e}^{\mathrm{j}2\pi k t_\Delta l \frac{T_s}{T_u}} X_{l,k} \mathrm{sinc}(\pi k t_\Delta) H_{l,k} + W_{l,k} + N_{t_\Delta}(l,k)$$

式中，l 是 OFDM 符号索引；k 是子载波索引；T_s和 T_u分别是总的符号持续时间和有效符号持续时间；$W_{l,k}$是加性高斯白噪声；N_{t_Δ}是由于采样频率偏移造成的加性噪声。式中最后一项的功率可以近似为

$$P_{t_\Delta} \approx \frac{\pi^2}{3}(k t_\Delta)^2$$

因此，干扰功率与偏移 t_Δ 和子载波索引 k 乘积的二次方成正比，这也就意味着相隔越远的子载波受到的影响越大。干扰功率还可以用信噪比损失来表示：

$$D_n \approx 10\log_{10}\left[1+\frac{\pi^2}{3}\frac{E_s}{N_0}(kt_\Delta)^2\right]$$

在子载波数量和采样错误 t_Δ 很小（$kt_\Delta << 1$）的 OFDM 系统中，由采样频率误差造成的影响几乎可以忽略。由 $\{e^{j2\pi kt_\Delta l\frac{T_s}{T_u}}\}$ 项可以明显看出，最重要的问题是基于子载波下标 k 和符号下标 l 引起不同子载波上不同的旋转值。因此相隔最远的子载波，旋转的角度最大，此外旋转角度与符号下标 l 成正比。t_Δ 项是由定时环控制的，一般值都较小。但是随着 l 增加了旋转角度，该项也会增大，最终导致不能正确解调，从而导致 OFDM 接收机不能跟踪采样频率。采样偏差对 OFDM 系统信噪比的影响如图 1-10 所示。

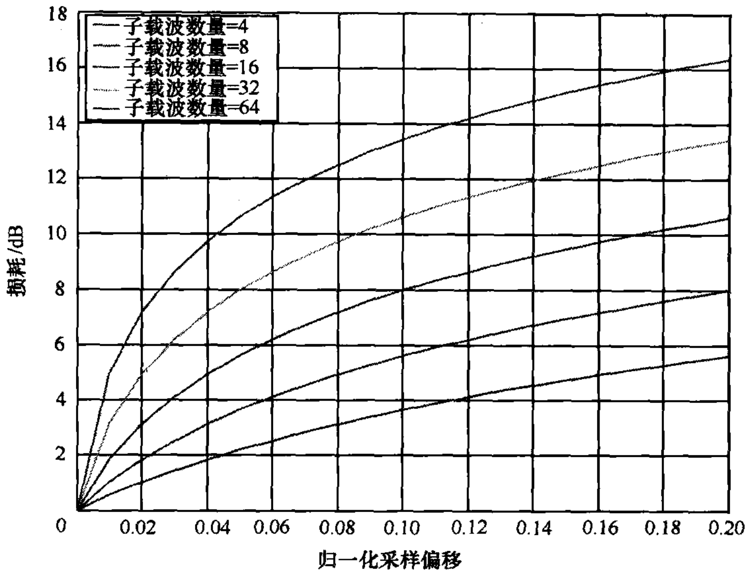


图 1-10 由于采样不匹配造成的信噪比下降

1.2.3.2 载频偏移

OFDM 系统与单载频系统相比更容易受到频率误差的影响。造成接收端频率偏移的主要原因是本地振荡器不稳定和发射机、接收机工作条件改变；发射机和接收机之间相对运动造成的多普勒频移；其他信道衰落造成的相位噪声等。所需子载波信号幅度降低和相邻子载波间的 ICI 会影响到系统性能。子载波信号的幅度衰减是由于所需的子载波不是在 DFT 等效 sinc 函数的峰值处被采样造成的。而相邻子载波的干扰是由于相邻子载波不是在过零点处被采样造成的。Pollet 等

在参考文献［6］中分析了载频偏移对信噪比的总影响，对于相对较小的载频偏移，信噪比（dB）的损失可以近似为

$$SNR_{\text{loss}} \approx \frac{10}{3\ln 10}(\pi T f_{\Delta})^2 \frac{E_s}{N_0}$$

式中，f_{Δ}是频率偏移，它与载波间隔和采样周期 T 有关。系统的性能还与调制类型有关。很显然，星座图点数多的调制方式相比星座图点数少的调制方式更容易受到载频偏移的影响，因为高阶星座图调制方案在同等误比特率（BER）性能下对信噪比的需求比低阶星座图调制方案要高。

假设 OFDM 系统的两个子载波可以用基带 A/D 转换器输出的正交频率分量表示为

$$\phi_k(t) = \mathrm{e}^{\mathrm{j}2fkt/T} \text{和 } \phi_{k+m}(t) = \mathrm{e}^{\mathrm{j}2\pi(k+m)t/T}$$

式中，T 是采样间隔。此外假设由于接收站的频率偏移 δ 造成第 k 个子载波对第（$k+m$）个子载波的频率偏移，也就是

$$\phi_{k+m}^{\delta}(t) = \mathrm{e}^{\mathrm{j}2\pi(k+m+\delta)t/T}$$

由于频率偏移的存在，第 k 个信道和第（$k+m$）个信道的干扰是

$$I_m(\delta) = \int_0^T \mathrm{e}^{\mathrm{j}k2\pi t/T}\mathrm{e}^{-\mathrm{j}(k+m+\delta)2\pi t/T}\mathrm{d}t = \frac{T(1-\mathrm{e}^{-\mathrm{j}2\pi\delta})}{\mathrm{j}2\pi(m+\delta)}$$

$$|I_m(\delta)| = \frac{T|\sin(\pi\delta)|}{\pi|m+\delta|}$$

所有 N 个子载波累计干扰造成的功率损耗可以近似估计为

$$\sum_m I_m^2(\delta) \approx (T\delta)^2 \sum_{m=1}^{N-1} \frac{1}{m^2} \approx (T\delta)^2 \frac{23}{14} \quad (N >> 1 \text{ 时})$$

1.2.3.3 定时偏移

符号定时对接收机十分重要，因为要准确地解调和译码接收到的数据序列必须要有精确的定时信息。通过 OFDM 系统中附加在数据符号之外的训练序列可以同步定时信息。由于信道引起发射机和接收机的定时偏移可能导致接收机不能完全恢复出发射信号的定时参考信息。定时偏移进一步会导致子载波相位旋转。定时偏移可以通过使用循环前缀来消除，如果定时偏移引起的信道响应比循环前缀长度短，子载波间可以保持正交性。定时偏移可以用信道的相位偏移表示，而相位偏移可以通过计算信道冲击响应来估计。当接收机与接收到的数据流不能在时间上同步，那么接收符号的信噪比将受到影响。定时偏移的影响可以用输出信噪比（SNR）与最优采样时间 T_{optimal}的关系表示为

$$\xi = \frac{\Lambda(\tau)}{\Lambda(0)}$$

式中，T_{optimal}是自相关函数，τ 是最佳采样时刻 T_{optimal} 和实际接收符号时间之间的延迟。参数 τ 可以看成是随机变量，因为 τ 是在噪声中估计出来的，一般称为定

时抖动。两种特殊的归一化自相关函数的信号（基带时限信号和带限信号）的数学形式可以表示为

$$\Lambda(\tau)=\left[1-\frac{|\tau|}{T_{\text{symbol}}}\right]$$

$$\Lambda(\tau)=\frac{1}{N}\left[\frac{\sin(\pi NW\tau)}{\sin(\pi W\tau)}\right]$$

式中，W 是带限信号的带宽。单载波系统最适合用带限信号来表示，而 OFDM（多载波）系统更适合用时间受限信号来描述。对于单载波系统，定时抖动表现为带限信号的参考噪声相位，OFDM 系统中，在传输数据的同时传输导频子载波以估计残留相位误差。

Paez-Borrallo 在参考文献［7］中分析了由于时间偏移导致正交性受到的影响，分析的结果量化了时间偏移对子载波间干扰和正交性的影响。我们假设两个连续符号之间的定时偏移表示为 τ，那么接收机的接收数据流可以表示为

$$X_i=c_0\int_{-T/2}^{-T/2+\tau}\phi_k(t)\phi_l^*(t-\tau)\,\mathrm{d}t+c_1\int_{-T/2+\tau}^{T/2}\phi_k(t)\phi_l^*(t-\tau)\,\mathrm{d}t$$

式中

$$\phi_k(t)=\mathrm{e}^{\mathrm{j}2\pi fkt/T}$$

令 $m=k-l$，那么接收符号的幅度可以表示为

$$|X_i|=\begin{cases}2T\left|\dfrac{\sin m\pi\dfrac{\tau}{T}}{m\pi}\right|, & c_0\neq c_1\\ 0, & c_0=c_1\end{cases}$$

如果 $\tau \ll T$，对于简单分析，可以进一步简化为

$$\frac{|X_i|}{T}\approx\frac{2m\pi\dfrac{\tau}{T}}{m\pi}=2\frac{\tau}{T}$$

在 $\tau \ll T$ 时，上式与 m 无关。

我们可以计算出平均干扰功率为

$$\mathrm{E}\left[\frac{|X_i|^2}{T^2}\right]=4\left(\frac{\tau}{T}\right)^2\times\frac{1}{2}+0\times\frac{1}{2}=2\left(\frac{\tau}{T}\right)^2$$

由于定时误差带来的 ICI 的损失（单位为 dB）可以计算为

$$ICI_{\text{dB}}=10\log_{10}\left[2\left(\frac{\tau}{T}\right)^2\right]$$

1.2.3.4 载波相位噪声

载波相位误差是由于发射机和接收机晶体振荡器的误差引起的。相位旋转可能是定时误差或者是时间选择性信道的载波相位偏差引起的。Pollet 等在参考文

献［8］中分析了载波相位噪声对系统性能的影响。载波相位噪声可以建模成维纳过程$\theta(t)$，其中$E\{\theta(t)\}=0$，$E[\{\theta(t_0+t)-\theta(t_0)\}^2]=4\pi\beta|t|$，式中$\beta$（单位为Hz）表示非稳态载波生成器的洛伦兹功率谱密度的单边带谱线带宽。载波相位噪声对系统信噪比的影响（也就是弥补该影响所需要增加的信噪比）可以近似为

$$D\approx\frac{11}{6\ln 10}\left(4\pi N\frac{\beta}{W}\right)\frac{E_s}{N_0}$$

式中，W和E_s/N_0是符号的带宽和信噪比。可以看出，系统信噪比随着子载波数量的增加而增大。

1.2.3.5 多径问题

在移动无线通信中，接收机收集到发射信号通过不同的路径传输得到的信号，这些信号可能是直接到达接收机，也可能是由于反射，甚至是附近的障碍物的衍射到达接收机。这些到达接收机的到达路径可能互相干扰，并影响承载信息的信号。多径对信道的影响包括延迟扩展、信号强度衰减和频谱的扩展。信道的时变特性可以建模成窄带平稳过程。由于接收端天线接收到许多反射而来的信号，根据中央极限定理，到达的信号分布可以建模成复数高斯随机过程。接收信号的包络可以分解成缓慢变化的慢变衰落和快变衰落。包络的平均幅值受到不同路径的干扰相位影响而快速下降，该过程称为衰落。多径主要用于描述接收机天线接收到多个承载信息的信号。这样的信道可以用统计的方法来建模，并用信道相关函数来描述其特性。基带传输信号可以精确地建模成如下窄带过程：

$$s(t)=x(t)\mathrm{e}^{-2\pi f_c t}$$

假设多径传播是高斯散射，信道可以通过时变传播延迟、损耗因子和多普勒频移来描述。信道的时变脉冲响应可以写成

$$c(\tau_n,t)=\sum_n \alpha_n(\tau_n,t)\mathrm{e}^{-j2\pi f_{D_n}\tau_n(t)}\delta[t-\tau_n(t)]$$

式中，$c(\tau_n,t)$是信道在t时刻对于在$t-\tau_n(t)$时刻的脉冲的响应；$\alpha_n(t)$是在第n条路径上接收信号的衰减因子；$\tau_n(t)$是第n条路径的传播延迟；f_{D_n}是在第n条路径上接收信号的多普勒频移。

多普勒频移是由于发射机和接收机相对移动造成的，可以表示为

$$f_{D_n}=\frac{v\cos(\theta_n)}{\lambda}$$

式中，v是发射机和接收机的相对速度；λ是载波的波长；θ_n是发射机和接收机的相位角度。

传输信号通过信道传输的输出是

$$z(t)=c(\tau_n,\ t)*s(t)$$
$$z(t)=\sum_n \alpha_n[\tau_n(t)]\mathrm{e}^{-\mathrm{j}2\pi(f_c+f_{D_n})\tau_n(t)}x(t-\tau_n(t))\mathrm{e}^{-\mathrm{j}2\pi f_c t}$$

式中

$$\delta(t-\tau_n(t))*x(t)=x(t-\tau_n(t))$$
$$\delta(t-\tau_n(t))*\mathrm{e}^{-\mathrm{j}2\pi f_c t}=\mathrm{e}^{-\mathrm{j}2\pi f_c(t-\tau_n(t))}$$
$$\beta_n=\alpha_n[\tau_n(t)]\mathrm{e}^{-\mathrm{j}2\pi(f_c+f_{D_n})\tau_n(t)}$$

$z(t)$ 可以表示为

$$z(t)=\sum_n \beta_n x(t-\tau_n(t))\mathrm{e}^{-\mathrm{j}2\pi f_c t}$$

式中，β_n是高斯随机过程。信道响应函数 $c(\tau_n,\ t)$ 的包络是瑞利分布函数，因为信道响应是高斯随机过程的和。瑞利衰落信道的概率密度函数是

$$f_z(z)=\frac{z}{\sigma^2}\mathrm{e}^{-\left(\frac{z^2}{2\sigma^2}\right)}$$

一般来说，没有直射（LOS）路径的信道（也就是只有散射路径）都是瑞利衰落信道。而有直射路径的信道一般可以用莱斯概率密度函数来表示

$$f_z(z)=\frac{z}{\sigma^2}I_0\left(\frac{z\eta}{\sigma^2}\right)\mathrm{e}^{-\left(\frac{z^2+\eta^2}{2\sigma^2}\right)}$$

式中，I_0是0阶修正贝塞尔函数；η 和 σ^2分别是直射路径的均值和方差。Proakis 在参考文献［9］中给出了 $c(\tau,\ t)$ 的自相关函数如下：

$$\Lambda_c(\tau,\Delta t)=\mathrm{E}\{c(\tau,t)c^*(\tau,t+\Delta t)\}$$

此外，可以通过发射很窄的脉冲，在接收端互相关接收信号与发射信号的共轭延迟，从而测量出信道的自相关函数。信道的平均功率可以令 $\Delta\tau=0$，也就是 $\Lambda_c(\tau,\ \Delta t)=\Lambda_c(\tau)$ 得到。这个量称为功率延迟分布或多径密度分布。τ 的取值范围，也就是让 $\Lambda_c(\tau)$ 非零的范围称为信号多径延迟扩展，用 τ_m表示。多径延迟扩展是信道相关带宽的倒数，也就是

$$B_m\approx\frac{1}{\tau_m}$$

信道相关带宽在通信系统中起到很重要的作用。如果通信系统的信号带宽与信道的相关带宽比要小，那么系统将经历平坦衰落（也称为非频率选择性衰落）。由于平坦衰落可以通过系统链路预算时增加功率余量来克服，因此平坦衰落的通信系统接收端信号处理很容易实现。反之，如果信号带宽比信道相关带宽要宽，系统将经历频率选择性衰落，接收机需要能准确地判决出所需要的信号。如果信道的统计量在超过一个符号间隔内保持平稳，则称为慢变衰落信道，反之，如果统计量在一个符号间隔内快速变化，则称为快变衰落信道。一般的无线宽带信道都用慢频率选择性衰落来描述。

1.2.3.6 符号间干扰（ISI）问题

式（1-1）中调制器的输出如下：

$$x(t)=\sum_{n=-\infty}^{\infty}\left[\sum_{k=0}^{N-1}d_{n,k}\phi_k(t-nT_d)\right]$$

式（1-1）可以用第 n 个 OFDM 符号的离散形式改写成

$$x_n(k)=\sum_{k=0}^{N-1}d_{n,k}\phi_k(t-nT_d)$$

式中，$\phi_k(t)=\mathrm{e}^{\mathrm{j}2\pi fkt/T}$。

对信道符号的第 n 个数据块 d_{nP}，d_{nP+1}，…，d_{nP+P-1}，第 i 个子载波信号可以表示为

$$x_n^i(k)=\sum_{k=0}^{N-1}d_{nP+i,k}\mathrm{e}^{\mathrm{j}\frac{2\pi}{N}l_ik}\qquad i=0,1,2,\cdots,P-1;\ P\text{是子载波数量}$$

式中，l_i 是长度为 N 的复指数时间下标，也就是 $0\leqslant l_i\leqslant N-1$。

这些子载波信号叠加组成第 n 个 OFDM 符号，即

$$x_n(k)\equiv\sum_{i=0}^{P-1}x_n^i(k)=\sum_{i=0}^{P-1}d_{nP+i}\mathrm{e}^{\mathrm{j}\frac{2\pi}{N}l_ik}\tag{1-5}$$

D/A 转换器输出的发射信号可以表示为

$$s(t)\equiv\sum_{n}\left[\sum_{k=0}^{L-1}x_n(k)\delta(t-(nL+k)T_d)\right]$$

式中，L 是长度超过 N（子信道数）的数据符号。由于数据序列长度 L 大于 N，只需要一部分接收到的 OFDM 符号就可以解调出子载波。剩下的 $Q=L-N$ 个码则不需要用于解调，这些符号可以用于保护间隔，增加循环前缀来消除 OFDM 系统中的 ICI 问题。在多径和加性噪声环境下，接收到的 OFDM 符号可以表示为

$$r_n(k)=\sum_{i=0}^{L-1}x_n(i)h(k-i)+\sum_{i=0}^{L-1}x_{n-1}(i)h(k+L-i)+v_n(k)\tag{1-6}$$

式中，第一项表示所需要的承载信息的信号通过多径环境，第二项则表示之前传输的符号带来的干扰。假设多径信道的长度 L_h 比 OFDM 符号的长度 L 要小得多。此外，假设信道是因果性（也就是符号间干扰只是由之前的符号带来的），多径信道和保护间隔一样长，也就是 $L_h\leqslant Q$，那么接收到的信号可以划分为两个时间间隔。第一个时间间隔包含所需要的信号和前面符号的干扰。第二个时间间隔只包含有承载信息的符号，用数学式表示为

$$r_n(k)=\begin{cases}\sum_{i=0}^{L-1}x_n(i)h(k-i)+\sum_{i=0}^{L-1}x_{n-1}(i)h(k+L-i)+v_n(k),0\leqslant k\leqslant Q-1\\ \sum_{i=0}^{L-1}x_n(i)h(k-i)+v_n(k),Q\leqslant k\leqslant L-1\end{cases}$$

(1-7)

接下来，我们将研究符号间干扰对性能的影响。符号间干扰是由于承载信息的脉冲在时间上弥散造成的，信息符号在时间上扩展导致部分能量延展到邻近的符号时隙内。奈奎斯特准则为在带限信道下用奈奎斯特采样频率从而避免符号间干扰传输指明了方向，也就是引入了迫零均衡的方法。迫零均衡器的复杂性取决于信道多径延迟的程度。接收机由于不能通过均衡来修正信道的影响，并且由于修正接收机结构中对噪声的增强效果，从而导致性能下降。在式（1-7）中的第二项表示了符号能量扩散到相邻符号间隔。符号间干扰的影响在时域和频域上都会得到体现。

OFDM 系统最大的特性之一就是 OFDM 系统具有很强的对抗多径延迟扩展以及符号间干扰的鲁棒性。这主要是由于 OFDM 系统把数据比特分为一些并行的子载波，从而增长了码周期，进而减少了符号间干扰。对抗多径延迟扩展的鲁棒性可以通过在发射码之间增加保护间隔来增强。保护间隔为前一个符号的多径信号在目前符号信息到达之前留出足够多的时间来传输。最有效的保护间隔是符号的循环扩展。符号的尾部添加在码的头部之前的保护间隔内，从而有效地保持了子载波间的正交性。使用循环扩展符号，可以让傅里叶变换（译码）所需的采样数据长度超过符号长度，从而更好地对抗信道多径和符号时间同步误差。

只要多径延迟小于保护间隔，对多径的信号强度没有严格的限制，甚至可以超过最短路径的信号强度，所有路径的信号能量在接收机的输入端叠加，由于快速傅里叶变换（FFT）使能量集中，所以在译码端得到的是所有路径的总能量。然而，如果延迟路径的能量足够小，那么信道对系统并不会带来太大的问题。实际上大多数情况下，因为路径延迟超过保护间隔一般是由于相距很远的障碍物反射造成的，而这种路径反射而来的信号能量在接收机天线上已经经过很大的衰减。

由于子载波间正交性被破坏而造成的载波间干扰是 OFDM 系统最大的问题。而频率偏移、相位偏移或者多径弥散都会影响子载波的正交性。图 1-11 表示了多径的影响，其中子载波 1 与符号积分区间对齐，而子载波 2 则出现了延迟。此时，由于快速傅里叶变换的周期数不是子载波 2 的周期数的整数倍，接收机将受到载波间干扰。通过利用设计用于对抗符号间干扰的保护间隔可以有效地抑制载波间干扰。发射机和接收机的频率偏移使得接收信号产生了残留频率错误。频率偏移的影响可以通过扩展式（1-7）来分析：

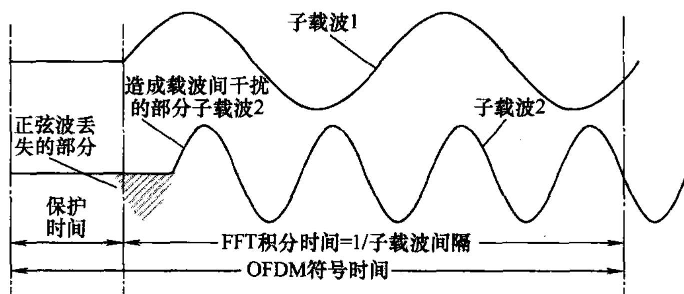


图1-11 多径对子载波间干扰的影响

$$r_n(k)=\begin{cases}\sum_{i=0}^{L-1}x_n(i)h(k-i)+\sum_{i=0}^{L-1}x_{n-1}(i)h(k+L-i)+v_n(k),\ 0\leqslant k\leqslant Q-1\\ \sum_{i=0}^{L-1}x_n(i)h(k-i)+v_n(k),\ Q\leqslant k\leqslant L-1\end{cases}\tag{1-8}$$

在接收机，去除保护间隔部分，剩下的信号部分 $k=0,\ 1\cdots N-1$ 可以定义为

$$r'_n(k)\equiv r_n(k+Q)\tag{1-9}$$

将式（1-5）代入式（1-9），简化后可以得到

$$r'_n(k)=\sum_{\alpha}h(\alpha)\sum_{i}d_{nP+i}\mathrm{e}^{\mathrm{j}\frac{2\pi}{N}l_i(k+Q-\alpha)}+v_n(k)$$

或

$$r'_n(k)=\sum_{i}d_{nP+i}\mathrm{e}^{\mathrm{j}\frac{2\pi}{N}l_ik}\mathrm{e}^{\mathrm{j}\frac{2\pi}{N}l_iQ}\sum_{\alpha}h(\alpha)\mathrm{e}^{-\mathrm{j}\frac{2\pi}{N}l_i\alpha}+v_n(k)\tag{1-10}$$

式（1-10）可以简化为

$$r'_n(k)=\sum_{i}d_{nP+i}\phi_iH(l_i)\mathrm{e}^{\mathrm{j}\frac{2\pi}{N}l_ik}+v_n(k)\tag{1-11}$$

不失一般性，在公式中省去（′）。

式中，ϕ_i为常数相位乘子，$\phi_i=\mathrm{e}^{\mathrm{j}\frac{2\pi}{N}l_iQ}$；$H(l_i)$ 为 $h(n)$ 的傅里叶变换，$H(l_i)=\sum_{\alpha}h(\alpha)\mathrm{e}^{-\mathrm{j}\frac{2\pi}{N}l_i\alpha}$。

将具有频率偏移 Δf 的接收信号代入式（1-11），可以得到

$$r_n^{\mathrm{off}}(k)\equiv r_n(k)\mathrm{e}^{\mathrm{j}2\pi\Delta fk}=\sum_{i}d_{nP+i}\phi_iH(l_i)\mathrm{e}^{\mathrm{j}\frac{2\pi}{N}k(l_i+\Delta fN)}+v_n(k)\tag{1-12}$$

从式（1-12）可以得出，频率偏移带来载波间干扰，并破坏了子载波间的正交性，从而影响了系统性能。换言之，符号估计为

$$\hat{d}_{nP+i}=G_i\left[\{H(l_i)d_{nP+i}I_{\Delta f}(0)\}+\left\{\sum_{\substack{i=0\\i\neq m}}^{P-1}H(l_m)d_{nP+i}I_{\Delta f}(l_m-l_i)\right\}+v_n(l_i)\right] \tag{1-13}$$

式中，载波间干扰项为

$$I_{\Delta f}(l_m-l_i)=\mathrm{e}^{\mathrm{j}\frac{2\pi}{N}k(l_m-l_i+\Delta fN)} \tag{1-14}$$

从式（1-14）中可以看出，由于较小的频率偏移带来的信噪比下降，近似为

$$SNR_{\text{loss}}\approx\frac{10}{3\ln10}(\pi\Delta fNT_s)^2\frac{E_s}{N_0} \tag{1-15}$$

式中，E_s/N_0是没有频率偏移时的信噪比。

回顾之前所述，通过在OFDM符号中引入保护间隔可以消除符号间干扰。保护间隔选择得比预计的延迟扩展要大，从而保护符号的多径部分不会影响其他相邻符号。这种保护间隔可以不包含信号，但是载波间干扰的问题仍然存在。如图1-12所示，OFDM符号在保护间隔中通过循环延展（有两种方法，即循环前缀或循环后缀）来消除载波间干扰。只要多径延迟小于保护间隔长度，就能保证经过多径的OFDM符号的延迟复制部分总是快速傅里叶变换周期的整数倍。通过循环延展，多径延迟小于保护间隔的信号经过多径并不会受到载波间干扰的影响。

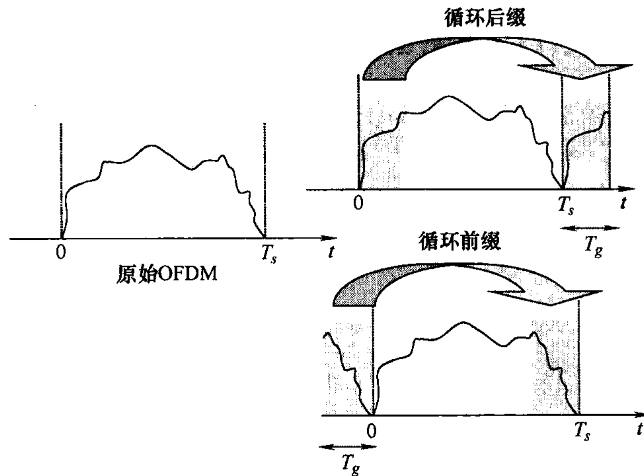


图1-12 保护间隔里的循环前缀

从数学上可以看出，OFDM符号在保护间隔中的循环延展，即使是在多径环境中造成信号延迟的情况下可以使OFDM符号在接收机保持周期性。在OFDM系统中，N个复数频域符号$X(n)$，$0<n<N-1$，通过离散傅里叶逆变换（IDFT）调制到N个正交子载波上的时域信号，可以表示为

$$x(k)=\sum_{n=0}^{N-1}X(n)\mathrm{e}^{+\mathrm{j}2\pi k\frac{n}{N}}=\mathrm{IDFT}\{X(n)\} \tag{1-16}$$

IDFT 的基函数是正交的，通过增加循环前缀，发射信号表现出周期性：

$$s(k)=\begin{cases}x(k+N), & 0\leqslant k<Q\\ x(k), & Q\leqslant k<L\end{cases}$$

式中，Q 是保护间隔的长度。接收信号可以改写成

$$y(k)=s(k)*h(k)+w(k),\ 0\leqslant k<L \tag{1-17}$$

如果增加的循环前缀比信号的脉冲响应长度长，信道的线性卷积在接收机端可以看成是圆周卷积，对任何子载波 l（$0\leqslant l<L$）有：

$$\begin{aligned}Y(n)&=\mathrm{DFT}(y(k))=\mathrm{DFT}(\mathrm{IDFT}(X(n))\otimes h(k)+w(k))\\&=X(n)\mathrm{DFT}(h(k))+\mathrm{DFT}(w(k))=X(n)H(n)+W(n),\ 0\leqslant k<N\end{aligned} \tag{1-18}$$

式中，$\otimes$表示圆周卷积；$W(n)=\mathrm{DFT}(w(k))$。式（1-18）表明子载波间不存在干扰。因此，通过增加循环前缀，在传输过程中保持了子载波间的正交性。使用循环前缀显然增加了需要发射的数据量，从而减少了可用的数据吞吐量。

1.2.3.7 功率峰均比（PAPR）

OFDM 系统（或多载波系统）的另一个问题是，由于信号的幅值动态范围过大，OFDM 信号的平均信号功率和信号峰值功率之间相差太大，从而造成峰均比过大。多载波调制的本质就是各子载波之间相互独立，因此这些子载波的幅值范围很大，子载波的幅值叠加或者抵消，从而导致总信号的功率变化范围很大。换言之，数据序列可能让所有的子载波的峰值对齐，从而使得信号功率很大，数据序列也可能让子载波的峰值相互抵消，从而产生功率很小的信号。信号功率变换范围太大就对发射机和接收机的设计带来很大的问题，因为发射机和接收机设计时要考虑到最小失真信号功率的范围。

OFDM 系统信号功率范围大，对于功率放大器（PA）和低噪声放大器（LNA）的设计带来很大的挑战。大的功率输出会把功率放大器带入非线性区域（也就是接近饱和），从而造成信号失真。为了减少失真，并且减少发射机的带外能量辐射总量，OFDM 和其他多载波调制都需要将功率放大器的工作区域限制在线性放大区内。由于 OFDM 本身较大的功率范围，OFDM 必须保持平均功率在功率放大器的非线性区域之下，以适应信号功率的变化。然而，降低平均功率会影响效率和功率范围，因为大多数信号需要用较低的输出功率传输，以抑制偶尔出现的功率峰值。因此，OFDM 设计时，必须在允许的信号失真和输出功率之间折中。也就是，OFDM 系统必须选择一个合适的平均输入功率，从而得到合适的输出功率，并且不会带来太多干扰或者违背功率谱限制。

为了进一步研究这种折中，考虑在 IEEE802.11a 中使用 52 个子载波的 OFDM 系统，理论上，52 个子载波可以峰值叠加，从而达到最大的峰值功率

比平均功率高 20log52dB = 34.4dB。然而，这是一个很少出现的情况。大多数仿真证明，实际的功率放大器足以实现平均功率高 3 ~ 6dB 的峰值功率，主要取决于功率放大器的性能和发射机的其他失真。换句话说，在功率放大器工作的允许功率范围以上的峰值带来的失真，大多数情况下不会降低平均错误率。

降低峰均比的一个简单办法是，通过消减幅值或者用较低幅值的平滑脉冲代替，以限制峰值信号。由于这种方法改变了信号的形状，因此某种程度上增大了信号失真。然而，如果控制信号幅值，则可以限制功率放大器带来的失真。因此该方法在很多情况下可以提高总输出功率效率。

对于基于分组的网络，接收机可以在分组传输出现错误时请求重传。通过扰码序列可以简单而有效地控制重传的峰均比。也就是，数据在调制到子载波上进行重传之前通过扰码，单独通过扰码不能避免出现大的峰值，在传输过程中仍然有可能由于分组数据中很大的峰值功率造成信号严重失真。在失真很严重时，接收机不能准确译码分组数据，因此接收机将请求重传。当重传数据时，由于使用的扰码序列不同，如果第一次传输使用的扰码造成较大的峰值功率，虽然重传的数据相同，但是第二次重传使用的扰码造成较大峰值功率的概率将会很低。由于 IEEE802.11a/g/n 网络已经使用了分组重传机制，该技术可以用于处理峰均比的一些问题。该技术的优点在于不会影响网络吞吐量，因为一些数据序列必须被重传。

很多具有不同复杂度和增强性能的技术已经提出用于减少由于峰均比对 OFDM 系统性能造成的影响。这些方案可以分为下面三类：

- 信号失真技术
  - 信号裁剪
  - 峰值窗
  - 峰值取消
- 编码技术
- 扰码技术

抑制峰均比过高问题最简单的方法是限制（裁剪）信号，使得信号的峰值总是在所允许的最大峰值功率以下。然而，这种办法会造成带外泄露和信号失真。这种裁剪的作用类似于对采样数据加上矩形窗，也等效于把信号的频谱与正弦函数（矩形窗函数的频谱）卷积，因此造成了频谱泄漏到旁瓣，从而干扰邻近的信道。所以为了抑制邻近信道的频谱泄漏，可以使用其他窄带宽的窗函数（例如高斯、凯撒、汉明和根升余弦函数）。

信号失真技术的目标是减少那些幅度超过特定门限的数据采样的幅值，由此带来的信号失真等问题可以通过使用峰值消除技术来避免。峰值消除技术使用从

信号中减去时间移位并缩放的参考信号，每个减去的参考信号都至少能减少一个信号采样的峰值功率，一般可以使用正弦函数作为参考信号，但是参考信号在频域上必须是带限的。正弦函数可以通过升余弦窗函数来实现频域带限。

研究者还考虑使用编码技术来抑制峰均比问题。为了实现更小的峰均比，就必须使用更低的编码速率。虽然有很多码组的性质可以用于前向纠错码（例如最小距离），但是这些码组并不适合于降低峰均比。然而，Wilkinson 等人在参考文献 [10] 中发现大多数可以实现降低峰均比的码组都是 Golay 互补序列，这种序列可以构建用于实现降低峰均比的编码。Golay 互补序列是一对自相关函数之和对于非零的延迟移位都是零的序列。

另一种直接的方法是扰码技术，对于任何 OFDM 符号，输入序列被一定数量的扰码序列加扰，然后选择具有最小峰均比的输出信号进行传输。

1.2.4 OFDMA

我们注意到 OFDM 传输在一个分组时间（由一些符号持续时间组成）内使用了所有的子载波。然而，我们可以引入更灵活的方法，让每个用户使用一部分频带（也就是一部分子载波），甚至部分分组时间（也就是一部分符号持续时间），这样多个用户可以共享 OFDM 传输，这也被称为正交频分复用接入（OFDMA）或多用户 OFDM。

我们已经把 OFDM 作为一种复用方案讨论，OFDM 提供了更高的频率效率和对抗无线环境多径衰落的能力。通过高度优化的 FFT/IFFT 模块，使得 OFDM 实现起来也很简单。这些优点同样可以扩展到多址接入方案，在多址接入时，子载波的子集分配给不同的用户。分配子载波子集给不同的用户，可以使得多个用户共享无线传输媒介，实现同时传输数据。从这个角度上来说，OFDMA 等效于普通的 FDMA 方式，然而 OFDMA 避免了 FDMA 方案中需要较大的保护带来分隔不同的用户。图 1-13 给出了一个 OFDMA 时频格的例子，其中 7 个用户 a ~ g 分别使用一部分可用子载波（每个用户使用的子载波部分可能不同）。空的时频格可以是未使用的，也可以用于传输导频信号。这个例子实际上是 OFDMA 和 TDMA 的混合，因为每个用户只在 4 个时隙中的一个内传输，而每个时隙可能包括一个或多个 OFDM 符号。

在前面的 OFDMA 例子中，每个用户拥有固定的子载波集。可以很容易想到把固定的子载波集合改进成允许每个时隙内在不同的子载波之间跳频的方案。允许每个用户使用不同的跳频模式，就相当于把 OFDMA 系统变成跳频 CDMA 系统。这样可以提高频率分集，因为每个用户可以使用所有的可用带宽，此外还有对于所有 CDMA 系统都具有的降低干扰的特点。通过对多次跳频使用前向纠错编码，系统可以对处于深衰落或受到其他用户干扰的子载波进行纠错。由于每次

频率

a		d		a		d		a		d	
a		d		a		d		a		d	
a	c	e		a	c	e		a	c	e	
a	c	e		a	c	e		a	c	e	
b		e	g	b		e	g	b		e	g
b		e	g	b		e	g	b		e	g
b		f	g	b		f	g	b		f	g
b		f	g	b		f	g	b		f	g

时间

图1-13 用户a、b、c、d、e、f、g共享时频格的OFDMA

跳频时干扰和衰落特性都在改变，系统性能取决于平均接收信号功率和干扰，而不是最差的衰落和干扰功率。跳频CDMA系统与直接序列扩频或多载波CDMA系统相比，其主要优点是通过使用小区内的正交跳频模式，能很方便地消除小区内干扰。

OFDMA在很多无线通信系统中得到了应用，例如IEEE 802.16e（OFDMA版本，也称为移动WiMAX）、IEEE 802.16m、3GPP长期演进（LTE）等。跳频OFDM也在超宽带（UWB）通信（也就是WiMedia）系统中使用，WiMedia即将成为蓝牙3.0的下一阶段系统。

1.2.4.1 无线资源分配

多用户OFDM（OFDMA）与OFDM最主要的区别在于无线资源分配，无线资源分配是一个典型的网络层问题，但是它与物理层传输密切相关。由于多个用户在时域（比特）和频域（子载波）共享OFDM传输，无线资源分配算法动态地优化利用时频资源格，也可以看成实际的在符号间干扰信道上执行注水算法，以达到香农容量限。

我们假设一共有K个用户，其中第k个用户数据速率为每个OFDM符号R_k比特。根据每个子载波上分配的比特数，自适应调制选择相应的调制方式，并根据子载波、比特和功率分配算法调整发射功率。我们定义$c_{k,n}$为第k个用户分配到第n个子载波上的比特数。为了建模这个问题，我们不允许超过一个用户共享一个子载波或者任何符号/比特。也就是，如果$c_{k',n}\neq 0$，$c_{k,n}=0\ \forall k\neq k'$，当$c_{k,n}\in\{0,1,\cdots,M\}=D$，$M$是每个OFDM符号的任意一个子载波上最多传输的比特数。我们用$f_k(c)$表示一个子载波上当信号增益归一化时可靠接收c个信息比特/符号所需的功率。为了在接收机获得所需的服务质量（QoS），第k个用户在第n个子载波上分配的传输功率等于

$$P_{k,n}=\frac{f_k(c_{k,n})}{\alpha_{k,n}^2}$$

我们的目标是找到最佳的 $c_{k,n}$ 的分配方案，从而使得总的发射功率在给定用户传输速率和给定 QoS 的限制下最小。此外，$f_k(c)$ 需要是凸的递增函数，并且 $f_k(0)=0$，这样就能保证符合实际的调制和编码方案。用数学式表示为

$$P_T^*=\min_{c_{k,n}\in D}\sum_{n=1}^{N}\sum_{k=1}^{K}\frac{f_k(c_{k,n})}{\alpha_{k,n}^2}$$

受到下列约束限制

约束1：$\forall k\in\{1,2,\cdots,K\}$，$R_k=\sum_{n=1}^{N}c_{k,n}$

约束2：$\forall n\in\{1,2,\cdots,N\}$，如果存在 k'，$c_{k',n}\neq 0$，那么 $c_{k,n}=0$，$\forall k\neq k'$

1.2.4.2 多用户 OFDM 的时间和频率同步

准确的解调 OFDM 信号需要保证子载波的正交性，因此需要时域和频域同步才能保持子载波的正交性。假设采样时钟同步准确（也就是在时间上整数倍采样持续时间对齐），我们试图估计出多用户 OFDM 系统的时间和频率偏移。

循环前缀带来的统计冗余可以提供时域和频域偏移的信息。

对于基站接收到的一个符号，我们假设 N 个子载波构成一个符号，N 个子载波又可以分为 M 组子载波（表示为 M_m 集合）。一个在第 m 个子载波组中传输的 OFDM 符号可以表示为

$$S_m(t)=\sum_{n\in M_m}X_n e^{j2\pi nt/(NT)}\qquad T_g<t<NT$$

式中，NT 是不考虑循环前缀一个 OFDM 符号的持续时间；T_g 是循环前缀长度；θ_m 是相对于接收机解调频率的频率偏移；ε_m 是相对于接收机符号时钟的时间偏移。

接收的信号是

$$\begin{aligned}r(k)&=\sum_{m=0}^{M-1}r_m(k)\\&=\sum_{m=0}^{M-1}s_m(k-m)e^{j2\pi\theta_m k/N}+n_m(k)\end{aligned}$$

式中，$s_m(k)$ 是发射信号。

如果估计器很准确，我们可以用带通滤波器来分隔子载波组。对第 m 个滤波器，我们可以获得一次性估计：

$$\hat{\varepsilon}_m=\arg\max_{\varepsilon}\{|r_m(\varepsilon)|-\rho_m R_m(\varepsilon)\}\tag{1-19}$$

$$\hat{\theta}_m=\frac{-1}{2\pi}\angle r_m(\varepsilon_m)$$

式中

$$r_m(\varepsilon) = \sum_{k=\varepsilon}^{\varepsilon+L-1} r_m^*(k) r_m(k+N)$$

$$R_m(\varepsilon) = \frac{1}{2}\sum_{k=\varepsilon}^{\varepsilon+L-1} |r_m(k)|^2 + |r_m(k+N)|^2$$

$$\rho_m = \frac{SNR_m}{SNR_m+1} \qquad SNR_m = \frac{\sigma_{S_m}^2}{\sigma_{n_m}^2}$$

如果每个滤波器的输出满足下式，那么这种估计器是时域偏移 ε 和频域偏移 θ 的联合最大似然（ML）估计。

$$r(k) = \tilde{S}(k-\varepsilon)e^{j2\pi\theta k/N} + n(k) \tag{1-20}$$

式中，$\tilde{S}(k)$ 是高斯分布，并且对于除了循环前缀的相似采样对之外，一般是不相关的。

1.3 MIMO

物理层通信的一个主要设计上的挑战在于如何提高频率效率，以提供在严重衰落信道上的宽带无线通信。从任何一本通信基础的书籍中都能找到相关的内容，例如参考文献［9，11］中提出对抗衰落的最佳办法之一就是利用分集来创建独立的通信信道。其中显而易见的一种方法是通过天线实现分集，也就是通过多个接收天线合并（最佳的合并方式是最大比合并（MRC））实现接收分集。Alamouti 在参考文献［13］中首先提出了两个发射天线最优的发射分集和接收分集，多输入多输出（MIMO）同时使用发射分集和接收分集来实现多个通信信道。在本节中，将介绍三类基础 MIMO 技术：

- 波束成形
- 空间复用
- 空时码

根据天线阵列的部署尺寸，可以分为两种基本的多天线方式：天线阵列相距很近的波束成形方式（天线间隔最多为 $\lambda/2$，其中 λ 是载波波长）；天线阵列相距较远的分集方式（典型的天线间隔至少是 λ 的几倍）。在本章中，我们将讨论后一种方法，只要天线间隔较大，两个传输-接收天线对的衰落过程可以认为是独立的。MIMO 系统由一些不相关的并行信道组成，可以从两个不同的方面来看，首先从分集的角度来看，信号的多个副本经过独立的衰落信道可以增强接收信号对抗衰落的能力。通过在并行独立衰落信道上发送同样的数据信号，可以有效地降低多径衰落的影响，并且降低中断概率，从而提高通信链路的可靠性。

第二种方法也称之为空间复用，不同的信息流在与发射天线对应的并行空间信道上传输。这可以看成是一种增加频谱效率的有效方法。为了能分离独立的数据流，一般来说，接收机必须装有数量多于并行信道数量的接收天线。对于给定的多天线配置，人们很想从上述两种方法中找到哪种能提供最优的性能。

1.3.1 空时码

空时码（STC）是一种综合利用了空间分集和时间分集的混合技术。有两种形式的空时码：空时分组码（STBC）和空时网格码（STTC）。STBC 在保持译码复杂度最小的前提下有效地利用发射分集对抗多径衰落。Tarokh 等在参考文献［19］中发现没有一种 STBC 可以在超过两根发射天线的情况下达到全速率和全分集，并且提出了四根发射天线的 3/4 速率全分集空时分组码。Jafarkhani 在参考文献［14］中根据 Alamouti 的正交 STBC 提出了四根发射天线全速率准正交（QO）STBC。此时，传输矩阵是

$$\mathbf{C}_j=\begin{bmatrix}\boldsymbol{A}_{12} & \boldsymbol{A}_{34}\\ -\boldsymbol{A}_{34}^* & \boldsymbol{A}_{12}^*\end{bmatrix}=\begin{bmatrix}x_1 & x_2 & x_3 & x_4\\ -x_2^* & x_1^* & -x_4^* & x_3^*\\ -x_3^* & -x_4^* & x_1^* & x_2^*\\ x_4 & -x_3 & -x_2^* & x_1\end{bmatrix} \tag{1-21}$$

式中，$\boldsymbol{A}_{12}$、$\boldsymbol{A}_{34}$是 Alamouti 码，注意，其中由于 QO-STBC 的信道矩阵是非满秩的，因此不能达到全分集增益。为了实现全分集和全速率（FDFR）特性，最近提出了新的 FDFR STC 构造方法。

混合模式（混合分集和空间复用）通过四根发射天线合并了发射分集和空间复用，每个空时分组用二阶基础 Alamouti 方案编码。第 i 个数据流（$i=a$，b）的空时分组编码传输矩阵为

$$\boldsymbol{A}_i=\begin{bmatrix}x_1(i) & x_2(i)\\ -x_2(i)^* & x_1(i)^*\end{bmatrix} \tag{1-22}$$

为了译码数据，可以使用最小均方差（MMSE）和迫零（ZF）接收机。对 MMSE 接收机，我们假设传输矩阵是［$a_{2n}(k)$，$a_{2n+1}(k)$，$b_{2n}(k)$，$b_{2n+1}(k)$］^{T}，其中 a 和 b 表示不同的信号流。首先确定抽头权重矢量和译码层数。如果第一个译码层是 a，该过程可以表示为

$$\begin{bmatrix}\hat{a}_{2n}(k)\\ \hat{a}_{2n+1}(k)\end{bmatrix}=\text{decision}\left\{\begin{bmatrix}\boldsymbol{w}_1^{\mathrm{H}}(k)\\ \boldsymbol{w}_2^{\mathrm{H}}(k)\end{bmatrix}\boldsymbol{y}_n(k)\right\} \tag{1-23}$$

可以使用 $\hat{a}_{2n}(k)$ 和 $\hat{a}_{2n+1}(k)$ 来消除原来信号的干扰，相应地，其他信号流可以根据下面译码：

$$\boldsymbol{y}'_n(k)=\boldsymbol{y}_n(k)-[\boldsymbol{h}_1(k)\boldsymbol{h}_2(k)]\begin{bmatrix}\hat{a}_{2n}(k)\\ \hat{a}_{2n+1}(k)\end{bmatrix}$$

$$\begin{bmatrix}\hat{b}_{2n}(k)\\ \hat{b}_{2n+1}(k)\end{bmatrix}=\text{decision}\left\{\begin{bmatrix}\boldsymbol{h}_3(k)\\ \boldsymbol{h}_4(k)\end{bmatrix}\boldsymbol{y}'_n(k)\right\} \tag{1-24}$$

为了便于对比，我们还使用了最大似然（ML）译码（见 1.3.1.1 节）来获得最优译码性能作为参考基线。

下面，我们简要介绍空间复用方案。V-BLAST（贝尔实验室垂直分层空时）架构最近因为能在复杂散射的无线环境中实现较高的频谱效率而被提出。在这种方法中，一种检测方式是使用传统的自适应天线阵列技术，也就是线性合并和调零。从概念上说，每个数据流（也就是每层）轮流被认为是要检测的信号，而其他信号当作干扰考虑。调零是通过对接收信号线性加权以满足一些性能相关的准则来实现的，例如 ZF 或 MMSE。线性调零方式是可行的，但是如果使用非线性调零将可以达到更好的性能。一种特殊的非线性调零技术是利用符号抵消以及线性调零进行信号的检测。通过符号抵消技术，可以从接收的信号矢量中消去已检测的信号部分对检测的干扰，从而有效地降低总体干扰。在这里，我们只考虑具有排序串行干扰消除的 ZF 和 MMSE 接收机。同样地，ML 译码接收机也用作参考接收机。

我们假设 $H_{ij}(k)$ 是从第 j 个发射天线到第 i 个接收天线的信道系数，w 是白高斯噪声，其协方差矩阵 $\boldsymbol{C}_w=\mathrm{E}[\boldsymbol{w}\boldsymbol{w}^{\mathrm{H}}]=\sigma^2\boldsymbol{IR}$。因此，接收信号矩阵可以表示为

$$\boldsymbol{y}_n(k)=\boldsymbol{H}(k)\boldsymbol{x}_n(k)+\boldsymbol{w}(k) \tag{1-25}$$

式中，k 表示第 k 个子载波；$\boldsymbol{y}_n(k)=[y_1(k)\cdots y_{N_R}(k)]^{\mathrm{T}}$；$\boldsymbol{x}_n(k)=[x_1(k)\cdots x_{N_R}(k)]^{\mathrm{T}}$；$\boldsymbol{w}(k)$ 是 $(N_R\times 1)$ 维的噪声矢量。

1.3.1.1 最大似然译码（最优方案）

最大似然检测 $\boldsymbol{x}(k)$ 是通过最大化条件概率密度函数，等效于最小化对数似然函数：

$$\hat{\boldsymbol{x}}(k)=\min_{\boldsymbol{x}(k)}\{\boldsymbol{y}(k)-\boldsymbol{H}\boldsymbol{x}(k)\}^{\mathrm{H}}\{\boldsymbol{y}(k)-\boldsymbol{H}\boldsymbol{x}(k)\} \tag{1-26}$$

式中，$\boldsymbol{x}(k)\in$ 所有的可能星座集合。

众所周知，最大似然译码的复杂度很高，因此提出了下面能够实现次优性能，但是降低了实现复杂度的方案。

1.3.1.2 排序串行干扰消除（OSIC）

线性检测技术可以用于取代最大似然译码方法，例如 ZF 和 MMSE。为了改

进线性检测技术，我们可以根据接收信号强度译码，从接收的信号中提取出译码的信号。这个方法根据发射信号结构不同，称为 D-BLAST（贝尔实验室对角分层空时）和 V-BLAST。为了简化，我们考虑 OSIC。OSIC 的接收过程可以归纳成下列步骤：

步骤 1：计算抽头权重矩阵 $\boldsymbol{W}$。

步骤 2：找到具有最大信噪比的译码层。

步骤 3：检测：

$z_k(n) = \boldsymbol{W}_k^{\mathrm{H}} \boldsymbol{y}(n)$

$\hat{\boldsymbol{x}}_k(n) = \mathrm{decision}[z_k(n)]$

步骤 4：干扰消除：

$\boldsymbol{y}(n) = \boldsymbol{y}(n) - \hat{\boldsymbol{h}}_k(n)$

$\boldsymbol{H} = [\boldsymbol{h}_1, \cdots, \boldsymbol{h}_{k-1}, 0, \boldsymbol{h}_{k+1}, \cdots, \boldsymbol{h}_T]$

步骤 5：重复步骤 1，直到检测完所有的符号。

迫零（ZF）：代价函数可以表示为

$$J_{\mathrm{ZF}} = \{\boldsymbol{y}(k) - \boldsymbol{H}\hat{\boldsymbol{x}}(k)\}^{\mathrm{H}}\{\boldsymbol{y}(k) - \boldsymbol{H}\hat{\boldsymbol{x}}(k)\} \tag{1-27}$$

由于 J_{ZF} 是 $\hat{\boldsymbol{x}}(k)$ 的凸函数，$\hat{\boldsymbol{x}}(k)$ 可以使用最小极限来确定。因此，抽头权重矢量是

$$\boldsymbol{W} = \{\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}\}^{-1}\boldsymbol{H}^{\mathrm{H}} \tag{1-28}$$

最小均方差（MMSE）：考虑到噪声方差，代价函数可以表示为

$$J_{\mathrm{MMSE}} = E[\{\boldsymbol{y}(k) - \boldsymbol{H}\hat{\boldsymbol{x}}(k)\}^{\mathrm{H}}\{\boldsymbol{y}(k) - \boldsymbol{H}\hat{\boldsymbol{x}}(k)\}] \tag{1-29}$$

使用与迫零检测类似的方法，抽头权重矢量为

$$\boldsymbol{W} = \{\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H} + \sigma^2\boldsymbol{I}\}^{-1}\boldsymbol{H}^{\mathrm{H}} \tag{1-30}$$

注意，为了使用 MMSE 方式，必须事先估计噪声方差。

1.3.2 基于自适应多天线技术的空间复用

一些研究者已经考虑到了空间分集复用问题。Zheng 和 Tse 在参考文献 [15] 中研究了分集和空间复用的折中方案。Heath 和 Paulraj 在参考文献 [16] 中提出了分集和空间复用切换的方案。后者考虑了在固定速率系统中，接收机根据最大最小接收星座图欧氏距离的原则自适应地从分集和空间复用中选择一种传输方案，接收机通过一个 1 比特反馈信道通知发射机其选择的方案。为了确保固定比特速率，分集方案使用比空间调制方案更高阶的调制。Skjevling 等在参考文献 [17] 中提出了融合分集和空间复用的混合方案。该方案最优地分配天线到给定的固定传输组合（融合了分集和空间复用）。天线选择是根据完全信道反馈或者长期统计做出的。Gorokhov 等在参考文献 [18] 中研究了在天线子集选择时

复用增益和分集增益的关系，从而扩展了 Zheng 和 Tse 在参考文献［15］中的结论。

1.3.3 开环 MIMO 方案

Alamouti 在参考文献［13］中提出了著名的两根发射天线的正交全分集和全速率（FDFR）码，此时接收机需要有一个简单的线性译码器。Tarohk 等在参考文献［19］中证明了只存在两根发射天线的 FDFR 正交码，并且提出了一些在多于两根发射天线的空时分组码，这些空时码虽然不满足全速率，但是可以达到全分集。Jafarkhani 提出了准正交全速率码，但是不能达到全分集。Xin 等在参考文献［20］中和 Ma 等在参考文献［21］中，根据空时星座旋转提出了适合任意发射天线数的 FDFR 编码器。在发射天线数量为偶数时，Jung 等在参考文献［22］中获得了通过把 Alamouti 方案用星座旋转技术串行连接起来获得全分集全速率空时分组码，并且获得了编码增益。虽然基于 Alamouti 的空时星座旋转编码器（A-ST-CR）可以有效地实现全分集和全速率，但是编码复杂度过高仍然是一个实际实现中存在的问题，尤其是发射天线数量较少时，如 4 根发射天线。这主要是由于最大似然（ML）译码算法计算复杂度过高。

考虑到复杂度问题，本节考虑用新的旋转预编码方案串行连接二阶 Alamouti 码，从而进一步扩展了研究结果。通过在实际的星座图旋转操作之后合适地进行打孔和移位，编码过程可以很容易地分解成低阶矩阵空间的旋转操作。对旋转编码进行打孔和移位对接收机十分重要，因为通过打孔和移位使得信号解耦合，从而极大地降低最大似然译码复杂度。在本节中，提出的方法可以达到与传统方案同样的性能，同时极大地降低复杂度。

研究者使用基于范德蒙德矩阵的预编码来获得 FDFR 系统。把接收到的信号 x 乘以范德蒙德矩阵，矢量 $\boldsymbol{r}$ 的每个元素合并了所有的符号，正如下面基本的预编码公式：

$$\boldsymbol{r}=\boldsymbol{\Theta x}=\frac{1}{\sqrt{4}}\begin{bmatrix}1 & \alpha_0^1 & \alpha_0^2 & \alpha_0^3\\ 1 & \alpha_1^1 & \alpha_1^2 & \alpha_1^3\\ 1 & \alpha_2^1 & \alpha_2^2 & \alpha_2^3\\ 1 & \alpha_3^1 & \alpha_3^2 & \alpha_3^3\end{bmatrix}\begin{bmatrix}x_1\\ x_2\\ x_3\\ x_4\end{bmatrix}=\begin{bmatrix}r_1\\ r_2\\ r_3\\ r_4\end{bmatrix} \tag{1-31}$$

式中，$\alpha_i=\exp(\mathrm{j}2\pi(i+1/4)/N)$，$i=0, 1, \cdots, N-1$。

Xin 等在参考文献［20］和 Ma 等在参考文献［21］中在信息符号乘上范德蒙德矩阵之后使用了对角信道矩阵。这种线性预编码称为星座旋转操作。注意，虽然该方案成功地实现 FDFR，但是并没有最优化编码增益。Jung 在参考文献

[22] 中通过把星座旋转预编码的基本 Alamouti 方案串行连接起来，从而改进了编码增益，此时使用的传输信号为

$$S=\begin{bmatrix} r_1 & r_2 & 0 & 0 \\ -r_2^* & r_1^* & 0 & 0 \\ 0 & 0 & r_3 & r_4 \\ 0 & 0 & -r_4^* & r_3^* \end{bmatrix} \tag{1-32}$$

在接收端，信号可以表示为

$$y=\begin{bmatrix} y_1 \\ y_2^* \\ y_3 \\ y_4^* \end{bmatrix}=\frac{1}{\sqrt{2}}\begin{bmatrix} h_1 & h_2 & 0 & 0 \\ h_2^* & -h_1^* & 0 & 0 \\ 0 & 0 & h_3 & h_4 \\ 0 & 0 & h_4^* & -h_3^* \end{bmatrix}\begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \end{bmatrix}+\begin{bmatrix} n_1 \\ n_2^* \\ n_3 \\ n_4^* \end{bmatrix}=\boldsymbol{Hr}+\boldsymbol{n} \tag{1-33}$$

由于 r_1、r_2、r_3、r_4 通过范德蒙德矩阵在 $x_1 \sim x_4$ 上已经加过，每个符号两次经过信道。我们可以把 r_1、r_2、r_3、r_4 分成两部分：（r_1、r_3 和 r_2、r_4）或（r_1、r_4 和 r_2、r_3），因此预编码的范德蒙德矩阵不必为 4×4 矩阵，也可以是更小的矩阵。因此，我们可以在星座旋转过程之后使用打孔和移位操作，从而形成新的预编码器：

$$\boldsymbol{r}_{1,3}=\begin{bmatrix} r_1 \\ r_3 \end{bmatrix}=\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & \alpha_0^1 \\ 1 & \alpha_2^1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_3 \end{bmatrix} \qquad \boldsymbol{r}_{1,4}=\begin{bmatrix} r_1 \\ r_4 \end{bmatrix}=\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & \alpha_0^1 \\ 1 & \alpha_3^1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_4 \end{bmatrix}$$
$$\text{或} \tag{1-34}$$
$$\boldsymbol{r}_{2,4}=\begin{bmatrix} r_2 \\ r_4 \end{bmatrix}=\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & \alpha_1^1 \\ 1 & \alpha_3^1 \end{bmatrix}\begin{bmatrix} x_2 \\ x_4 \end{bmatrix} \qquad \boldsymbol{r}_{2,3}=\begin{bmatrix} r_2 \\ r_3 \end{bmatrix}=\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & \alpha_1^1 \\ 1 & \alpha_2^1 \end{bmatrix}\begin{bmatrix} x_2 \\ x_3 \end{bmatrix}$$

在打孔和移位之后，编码器可以定义为

$$\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & \alpha_0^1 & 0 & 0 \\ 0 & 0 & 1 & \alpha_1^1 \\ 1 & \alpha_2^1 & 0 & 0 \\ 0 & 0 & 1 & \alpha_3^1 \end{bmatrix} \quad \text{或} \quad \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & \alpha_0^1 & 0 & 0 \\ 0 & 0 & 1 & \alpha_1^1 \\ 0 & 0 & 1 & \alpha_2^1 \\ 1 & \alpha_3^1 & 0 & 0 \end{bmatrix} \tag{1-35}$$

最近，Rajan 等提出了三根发射天线和四根发射天线配置下具有低译码复杂度的带全分集的空时编码。传输矩阵的格式修正为

$$s=\begin{bmatrix} x_1+\mathrm{j}y_3 & -x_2+\mathrm{j}y_4 & 0 & 0 \\ x_2+\mathrm{j}y_4 & x_1-\mathrm{j}y_3 & 0 & 0 \\ 0 & 0 & x_3+\mathrm{j}y_1 & -x_4+\mathrm{j}y_2 \\ 0 & 0 & x_4+\mathrm{j}y_2 & x_3-\mathrm{j}y_1 \end{bmatrix} \tag{1-36}$$

式中，$x_i=s_{iI}\cos\theta-s_{iQ}\sin\theta$；$y_i=s_{iI}\cos\theta-s_{iQ}\sin\theta$；$\theta=\arctan(1/3)$。复符号 s_i 取值

于正交调幅（QAM）信号集。

1.3.4 闭环 MIMO 方案

在本节中，我们将介绍广泛使用的闭环 MIMO 方案，它分为两个部分，也就是天线分组和使用移动台反馈信息的码本方案。

1.3.4.1 天线分组

在 IEEE 802.16e 中，四根发射天线基站使用的码率为 1 的传输空时码是

$$A=\begin{bmatrix} s_1 & -s_2^* & 0 & 0 \\ s_2 & s_1^* & 0 & 0 \\ 0 & 0 & s_3 & -s_4^* \\ 0 & 0 & s_4 & s_3^* \end{bmatrix} \tag{1-37}$$

注意，该方案不能达到全分集。使用等效模型

$$A^{\mathrm{H}}A=\begin{bmatrix} \rho_1 & 0 & 0 & 0 \\ 0 & \rho_1 & 0 & 0 \\ 0 & 0 & \rho_2 & 0 \\ 0 & 0 & 0 & \rho_2 \end{bmatrix} \tag{1-38}$$

如果基站可以使用信道状况信息，矩阵 A 的性能接近了全分集全速率 STC 的性能：

$$\arg \min_{\text{antenna_pair}} |\rho_1-\rho_2| \tag{1-39}$$

令 $d_{\min}^2$ 是与归一化能量星座图对应的最小距离。使用 2^RQAM 欧氏距离等式 $d_{\min}^2=12/(2^R-1)$ 对应用于分集的 QAM 方式。使用该欧氏距离等式，我们可以估计出错误概率为

$$P_e \leqslant N_e Q\left(\sqrt{\frac{E_s}{N_0}d_{\mathrm{MIN}}^2}\right) \tag{1-40}$$

式中，d_{MIN}^2是接收信号欧氏距离的二次方；N_e是星座图上最近的星座点数量，对每个可能的星座图映射方案，根据信道系数矩阵 $HQ(x)=\mathrm{erfc}(x/\sqrt{2})/2$ 可以得到 N_e（式中，erfc 是余误差函数）。对于空时码，接收机分集星座图的最小距离是

$$d_{\mathrm{MIN}}^2(\boldsymbol{H}) \leqslant \frac{\min(\|\boldsymbol{H}\|_F^2(a,b),\|\boldsymbol{H}\|_F^2(c,d))}{N_T}d_{\min}^2 \tag{1-41}$$

式中，$(a,\ b)$ 和 $(c,\ d)$ 是天线分组下标；$\|\boldsymbol{H}\|_F$是矩阵 $\boldsymbol{H}$ 的 Forbenius 范数。详细的推导过程是根据分组码设计的最大信噪比原则给出。图 1-14 给出了系统框图，该系统使用天线分组，根据移动台反馈的信道信息选择出天线对。

该方案的性能比没有使用天线分组的传统空时码高 3.5dB。

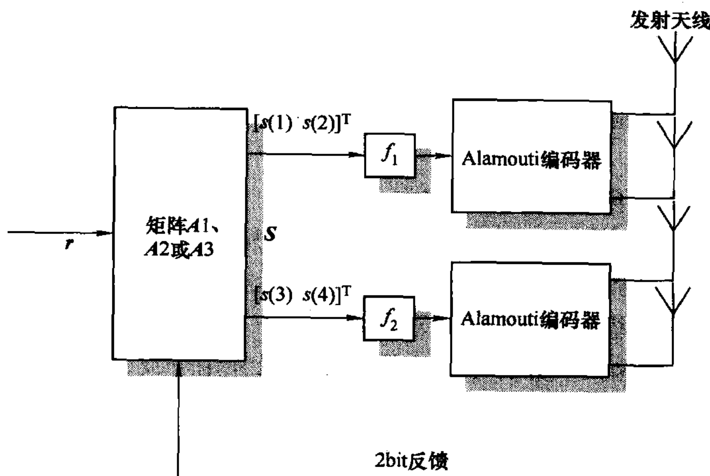


图 1-14 系统框图

目前 IEEE802.16e 标准中，四根传输天线的速率为 2 的传输空时码是

$$B=\begin{bmatrix} s_1 & -s_2^* & s_5 & -s_7^* \\ s_2 & s_1^* & s_6 & -s_8^* \\ s_3 & -s_4^* & s_7 & s_5^* \\ s_4 & s_3^* & s_8 & s_6^* \end{bmatrix} \tag{1-42}$$

在移动台中，最优的传输矩阵根据下列准则确定。令 $y_{r,i}$ 是第 i 个符号时间在第 r 根接收天线接收到的信号，$h_{i,r}$ 表示第 i 根发射天线和第 r 根接收天线之间的信道参数。当接收天线的数量为 2 时，接收的信号可以表示为

$$y=X(HW)s+v \tag{1-43}$$

式中，$y=[y_{1,1} \quad y_{1,2}^* \quad y_{2,1} \quad y_{2,2}^*]^T$；$s=[s_1 \quad s_2 \quad s_3 \quad s_4]^T$；$v$ 是噪声矢量；

$$H=\begin{bmatrix} h_{1,1} & h_{2,1} & h_{3,1} & h_{4,1} \\ h_{1,2} & h_{2,2} & h_{3,2} & h_{4,2} \end{bmatrix}$$

$X(\cdot)$ 是 2×4 的输入矩阵，定义为

$$X\left(\begin{bmatrix} a & b & c & d \\ e & f & g & h \end{bmatrix}\right)=\begin{bmatrix} a & b & c & d \\ b^* & -a^* & d^* & -c^* \\ e & f & g & h \\ f^* & -e^* & h^* & -g^* \end{bmatrix} \tag{1-44}$$

在移动台中，传输矩阵 B_q 的下标根据下列准则确定：

$$q=\arg\min_{l=1,2,\cdots,6}[\text{abs}(\det(H_{l,1})+\det(H_{l,2}))] \tag{1-45}$$

式中，$H_{l,1}$ 是矩阵 HW_l 的前两列；$H_{l,2}$ 是该矩阵最后两列。注意式（1-45）中的

天线分组矩阵选择规则等效于下列准则：

$$q = \arg\min_{l=1,2,\cdots,6}[\operatorname{trace}([(\boldsymbol{X}(\boldsymbol{HW}_l))^{\mathrm{H}}\boldsymbol{X}(\boldsymbol{HW}_l)]^{-1})] \tag{1-46}$$

天线分组的其他准则也可以用于确定天线分组下标。例如，最小误码率、最小均方差等。

1.3.4.2 **基于闭环 MIMO 的码本**

从移动台（MS）到基站（BS）的反馈中使用了码本。移动台从下行获得信道状态信息，并且从码本中选择出发射波束成形矩阵。然后 MS 把选择的码本中矩阵的序号反馈给基站。每个码本对应着一种 N_t、N_s和 N_i的组合，其中 N_t、N_s和 N_i分别是 BS 传输天线数、可用的数据流数量和用于反馈序号的比特的数量。一旦 MS 确定了 N_t、N_s和 N_i，MS 将在每 N_i个比特中反馈码本序号。在接收到 N_i个比特的序号之后，BS 将查找相应的码本，并根据序号找到 MS 选择的矩阵（或矢量）。选择的矩阵将用于 MIMO 预处理的波束成形矩阵。

1.3.5 MIMO 接收机结构

考虑未编码的 MIMO 传输为

$$\boldsymbol{y} = \boldsymbol{Hx} + \boldsymbol{n} \tag{1-47}$$

式中，$\boldsymbol{n}$ 是零均值和方差的空时循环对称复高斯白噪声矢量；$\boldsymbol{H}$ 是信道矩阵。

最大似然检测 $\boldsymbol{x}$ 得到

$$\begin{aligned}\hat{\boldsymbol{x}} &= \underset{\boldsymbol{x}}{\operatorname{argmin}}\|\boldsymbol{y} - \boldsymbol{Hx}\|^2 \\ &= \underset{\boldsymbol{x}}{\operatorname{argmin}}\sum_{i=1}^{M_R}\left|y_i - \sum_{j=1}^{M_T} h_{i,j}x_j\right|^2\end{aligned}$$

式中，$h_{i,j}$是矩阵 $\boldsymbol{H}$ 的第（i, j）个元素。最大似然检测算法的复杂性随着发射天线数量 M_T和 Q（Q 是调制中每个符号的比特数）的增长而呈指数增加。

1.3.5.1 **线性接收机**

基本思路是把接收到的信号通过线性变换进行预处理：

$$\tilde{\boldsymbol{y}} \equiv \boldsymbol{Ay} = \boldsymbol{AHx} + \boldsymbol{An}$$

因此，$\boldsymbol{AH}$ 近似于对角矩阵。选择预处理矩阵最直接的方法是使用迫零（ZF）接收机移除 $\boldsymbol{AH}$ 的非对角线元素。

$$\tilde{\boldsymbol{y}}_{\mathrm{ZF}} = \boldsymbol{H}^{+}\boldsymbol{y}$$

式中，$\boldsymbol{H}^{+}$是 $\boldsymbol{H}$ 的 Moore-Penrose 伪逆。

1.3.5.2 **LMMSE 接收机**

另一种方法是联合最小化 $\boldsymbol{AH}$ 的非对角线元素和经过过滤的噪声 $\boldsymbol{A}_n$，称之为线性最小均方差（LMMSE）：

$$\tilde{\boldsymbol{y}}_{\text{LMMSE}} = \left(\boldsymbol{H}^{\text{H}}\boldsymbol{H} + \frac{N_0}{E}\boldsymbol{I}\right)^{-1}\boldsymbol{H}^{\text{H}}\boldsymbol{y}$$

式中，E 是 $\boldsymbol{x}$ 中一个元素的平均能量；$\boldsymbol{I}$ 是 $\boldsymbol{M}_T \times \boldsymbol{M}_T$维的单位矩阵。

线性接收机一般具有较低的系统复杂性，但是性能较差，尤其是在 $\boldsymbol{M}_T = \boldsymbol{M}_R$时。

1.3.5.3 **判决反馈接收机**

判决反馈检测的预处理是把信道矩阵 $\boldsymbol{H}$ 分解成矩阵乘积的形式：

$$\boldsymbol{H} = \boldsymbol{QR}$$

为了简化，我们假设 $M_R \geqslant M_T$。那么 $\boldsymbol{Q}^{\text{H}}\boldsymbol{Q} = \boldsymbol{I}_{M_T}$，并且 $\boldsymbol{R}_{M_T \times M_T}$是上三角矩阵。矩阵分解后，变换的观察矢量 $\tilde{\boldsymbol{y}} \equiv \boldsymbol{Q}^{\text{H}}\boldsymbol{y}$ 的形式可以表示为

$$\tilde{\boldsymbol{y}} \equiv \boldsymbol{Rx} + \tilde{\boldsymbol{n}}$$

式中，$\tilde{\boldsymbol{n}} \equiv \boldsymbol{Q}^{\text{H}}\boldsymbol{n}$ 保持了 $\boldsymbol{n}$ 的统计特性。最小化 $m(\boldsymbol{x}) \equiv \|\boldsymbol{y} - \boldsymbol{Hx}\|_F^2$也就等效于最小化

$$\tilde{\boldsymbol{m}}(\boldsymbol{x}) \equiv \|\boldsymbol{y} - \boldsymbol{Rx}\|^2$$

从 $\boldsymbol{R}$ 的结构可以得到检测算法就是通过最小化 $|\tilde{y}_{M_T} - r_{M_T,M_T}\tilde{x}_{M_T}|^2$来检测 x_{M_T}，然后使用 x_{M_T}最小化 $|\tilde{y}_{M_T-1} - r_{M_T,-1,M_T-1} - r_{M_T,-1,M_T}\tilde{x}_{M_T}|^2 + |\tilde{y}_{M_T} - r_{M_T,M_T}\tilde{x}_{M_T}|^2$来检测 x_{M_T-1}。

很明显，该算法容易受到差错传播的影响。

1.3.5.4 **球形检测**

球形检测算法（SDA）可以以较低的复杂度达到接近最大似然检测的性能。基本思路是在一些小的候选集合中搜寻最优的 $\boldsymbol{x}$。一般地，该搜寻过程局限在以 y 为中心、r 为半径的超球面：

$$\|\boldsymbol{y} - \boldsymbol{Hx}\|^2 \leqslant r^2$$

如果 $r \to \infty$，那么球形检测就等效于最大似然检测。复杂度的降低主要是通过选择合适的半径 r 来实现的。

SDA 可以通过图 1-15 所示的树来实现，其中树的底部叶节点对应着所有可能的矢量 $\boldsymbol{x}$，其中 $\boldsymbol{x}$ 的元素代表着从底部到顶部的分支。

此外，$\tilde{\boldsymbol{y}} = \boldsymbol{Q}^{\text{H}}\boldsymbol{y}$，最大似然检测等效于通过对所有可能的分支计算尺度 $\tilde{m}(\boldsymbol{x}) \equiv \|\boldsymbol{y} - \boldsymbol{Rx}\|^2$，并获得最小值来遍历树。换句话说，SDA 通过合适地修剪树，减少了树的分支数。算法的过程可以归纳如下：

1）使用判决反馈接收机获得对 $\boldsymbol{x}$ 的初始估计 $\hat{\boldsymbol{x}}$，计算相应的尺度 $\tilde{m}(\hat{\boldsymbol{x}}) \equiv \|\hat{\boldsymbol{y}} - \boldsymbol{Rx}\|^2$。

2）该值设置成球的半径的二次方。然后深度优先，从顶到底遍历树，通过

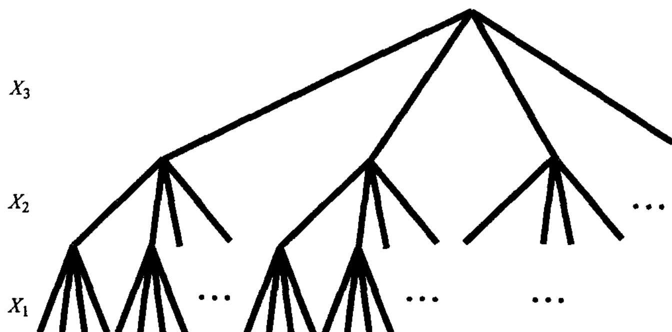


图 1-15 $M_T=3$，四进制调制的 MIMO 的 SDA

逐渐递增的方法计算尺度：

$$\|\hat{\boldsymbol{y}}-\boldsymbol{R}\boldsymbol{x}\|^2=\sum_{i=1}^{M_T}|\hat{y}_i-(\boldsymbol{R}\boldsymbol{x})_i|^2$$

3）当在某个节点的部分累加值大于或等于 $\tilde{m}(\tilde{\boldsymbol{x}})$ 时，该节点下不存在树叶（也就是不需要再考虑该节点）。

4）如果找到了比尺度 $\tilde{m}(\tilde{\boldsymbol{x}})$ 更小的新节点 $\boldsymbol{x}$，那么在接下去的算法中将新节点取代旧节点 $\boldsymbol{x}$（也就是减少了最大似然搜索中的球形空间范围）。

1.3.5.5 未编码信号的 MIMO 接收机

在 MIMO 信道的输出端检测未编码的信号，可以看成是一般信道进行均衡的特例。

考虑 M_T根发射天线、M_R根接收天线的 MIMO 信道。信道的输入输出关系可以用条件概率密度函数来表示：

$$f(\boldsymbol{y}|\boldsymbol{x})\propto \mathrm{e}^{\|\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x}\|_F^2/N_0}=\prod_{j=1}^{M_R}f(y_j|\boldsymbol{x})$$

式中，$f(y_j|x)=\mathrm{e}^{-|y_j-\boldsymbol{h}_j x|^2/N_0}$，$\boldsymbol{h}_j$是矩阵 $\boldsymbol{H}$ 的第 j 行。

$$f(\boldsymbol{y}|\boldsymbol{x})\propto\prod_{i=1}^{M_T}f(x_j)f(\boldsymbol{y}|x_1,\cdots,x_{M_T})=\prod_{i=1}^{M_T}f(x_i)\prod_{j=1}^{M_R}f(y_j|\boldsymbol{x})$$

为了降低迭代带来的复杂度，我们可以考虑下面的次优 MIMO 接收机：

1）对接收信号预处理以限制空间干扰：

$$\tilde{\boldsymbol{y}}=\boldsymbol{A}(\boldsymbol{H})\boldsymbol{y}=\boldsymbol{A}(\boldsymbol{H})\boldsymbol{H}\boldsymbol{x}+\boldsymbol{A}(\boldsymbol{H})\boldsymbol{n}$$

2）简单交换消息；

3）执行由有限步骤的单次扫描组成的检测。

回顾之前

$$\tilde{\boldsymbol{y}}_{ZF}=\boldsymbol{H}^{+}\boldsymbol{y}$$

加上噪声，可以得到：

$$\tilde{\boldsymbol{y}}=\boldsymbol{x}+\boldsymbol{H}^{+}\boldsymbol{n}$$

或者我们可以使用 LMMSE 以最小化 MSE：

$$E\{\|\boldsymbol{A}\boldsymbol{y}-\boldsymbol{x}\|_F^2\}$$

$$\boldsymbol{A}(\boldsymbol{H})=\left(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}+\frac{N_0}{E_s}\boldsymbol{I}\right)^{-1}\boldsymbol{H}^{\mathrm{H}}$$

由于 $\boldsymbol{A}(\boldsymbol{H})\boldsymbol{H}$ 的非对角线元素很小，检测算法可以根据以下估计：

$$f(\boldsymbol{y}_i\mid x_1,x_2,\cdots,x_{M_T})\approx f(\boldsymbol{y}_i\mid x_i)$$

在 M_R 远大于 M_T 时，ZF 和 LMMSE 接收机性能都很好。

1.3.5.6　非线性处理

根据下面操作，著名的 V-BLAST 被划分到次优的接收机：

- 空间干扰置零：通过修改预处理过程来达到。
- 空间干扰抵消：通过对 $i=M_T-1$，M_T-2，…，1 按

$$f(\tilde{y}_i\mid x_i,x_{i+1},\cdots,x_{M_T})\approx f(\tilde{y}_i\mid x_i,\tilde{x}_{i+1},\cdots,\tilde{x}_{M_T})$$

排序简化，式中 $\tilde{x}$ 表示根据 x 作出的判决。

- 排序：由于 V-BLAST 容易受错误传播的影响，天线排序会极大地影响接收机性能。

1.4　多用户检测（MUD）

实际上，最初的 MIMO 系统可以是图 1-16 所示的具有多个发射机和多个接收机的 CDMA 系统。然而，大多数 CDMA 系统使用了单用户接收机结构，并且把其他用户的信号当作同频干扰处理。Verdu 和 Poor 在参考文献 [23] 中提出了最优接收 CDMA 信号，之后多用户检测这一通信理论的新领域得到了极大的发展。

1.4.1　多用户 CDMA 接收机

CDMA 系统性能是由传统的相关接收机的多址干扰决定的，因此不能最优地检测信号。

每个用户以同样的数据速率 $1/T$ 传输数据信息。第 k 个用户的双极性数据流为

$$\boldsymbol{b}_k^{\mathrm{T}}=[b_k(-M),\cdots,b_k(0),\cdots,b_k(M)]$$

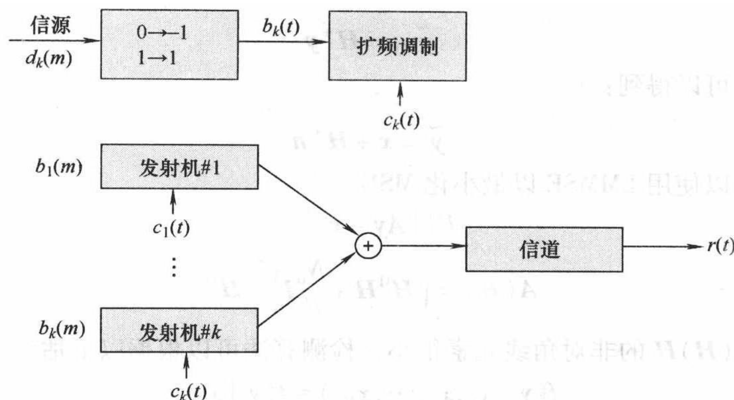


图 1-16 CDMA 系统

第 k 个用户的传输信号为

$$x_h(t)=\sum_{m=-M}^{M}b_k(m)c_k(t-mT-\tau_k)$$

式中，τ_k是第 k 个用户的时间延迟。

假设短码 $N=G_p=\frac{T}{T_c}$满足归一化条件，则

$$\frac{1}{T}\int_0^T c_h^2(t)\,\mathrm{d}t=1$$

接收的每个比特能量为

$$E_{c_k}\int_0^T b_k^2(m)c_k^2(t)\,\mathrm{d}t=E_{c_k}$$

因此，接收的波形为

$$r(t)=\sum_{k=1}^{K}\sqrt{E_{c_k}}\sum_{m=-M}^{M}b_k(m)c_k(t-mT-\tau_k)+n(t)$$

式中，$\tau_k<T$，在同步的情况下，$\tau_k=0$。

对于单个信号（$m=0$）

$$r(t)=\sum_{k=1}^{K}\sqrt{E_{c_k}}b_k(0)c_k(t)+n(t)$$

码的互相关为

$$\rho_{\ln}(j)=\frac{1}{T}\int_{T_l+mT}^{T_l+(m+1)T}c_l(t-\tau_l)c_m(t+jT-t_n)\,\mathrm{d}t$$

$$l=1,2,\cdots,K$$

$$n=1,2,\cdots,K$$

$$j=-M,\cdots,0,\cdots,M$$

传统的接收机使用单次判决，由于对信息的干扰不能忽略，因此传统接收机

达不到最优。在最优检测时，我们假设波形、延迟、相移、幅度和用户数都已知。

最优接收机可以检测出下列比特序列：

$$\hat{\boldsymbol{b}}=\begin{bmatrix}\hat{b}_1(-M) & \cdots & \hat{b}_1(M)\\ \vdots & & \vdots\\ \hat{b}_k(-M) & \cdots & \hat{b}_k(M)\end{bmatrix}$$

从而最大化条件概率

$$P[\hat{\boldsymbol{b}}\,|\,r(t)]$$

如果假设传输比特相互独立，并且等概率，也就等效于最大化

$$P[r(t)\,|\,\hat{\boldsymbol{b}}]=c\exp\left(-\frac{1}{2\sigma_n^2}\int_0^T[y_k(t)-\sum_{k=1}^{K}\hat{b}_k(0)\sqrt{E_{c_k}}c_k(t)]^2\mathrm{d}t\right)\qquad t\in[0,T]$$

通过维特比算法或者最大似然序列检测（见图 1-17）可以达到最优检测。但是由于检测算法是 NP 难题，因此其算法复杂度随着用户数量呈指数增长。

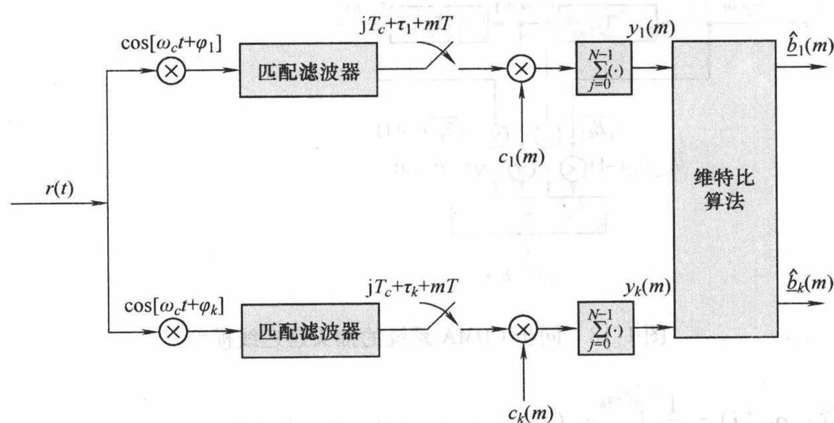


图 1-17　最佳接收机

我们考虑两用户情况：$T_1=0$，$T_2\leqslant T$。匹配滤波器第 m 个采样时刻的输出是

$$y_1(m)=\frac{1}{T}\int_{mT}^{(m+1)T}r(t)c_1(t)\mathrm{d}t$$

$$y_2(m)=\frac{1}{T}\int_{mT+T_2}^{(m+1)T+T_2}r(t)c_2(t-T_2)\mathrm{d}t$$

式中

$$r(t)=\sqrt{E_{c_1}}\sum_{i=-M}^{M}b_1(i)c_1(t-iT_b)+\sqrt{E_{c_2}}\sum_{i=-M}^{M}b_2(i)c_2(t-iT_b-\tau_2)$$

$$y_1(m)=\sqrt{E_{c_1}}b_1(m)+\sqrt{E_{c_2}}b_2(m-1)\rho_{12}(1)+\sqrt{E_{c_2}}b_2(m)\rho_{12}(0)+\sqrt{E_{c_2}}b_2(m+1)\rho_{12}(-1)+n_1(m)=z_1(\boldsymbol{b}_1,\boldsymbol{b}_2)+n_1(m)$$

$$y_2(m)=\sqrt{E_{c_2}}b_2(m)+\sqrt{E_{c_1}}b_1(m-1)\rho_{21}(1)+\sqrt{E_{c_1}}b_1(m)\rho_{21}(0)+\sqrt{E_{c_1}}b_1(m+1)\rho_{21}(-1)+n_2(m)=z_2(\boldsymbol{b}_1,\boldsymbol{b}_2)+n_2(m)$$

$$\boldsymbol{b}_1=[b_1(m-1),b_1(m),b_1(m+1)]$$

$$\boldsymbol{b}_2=[b_2(m-1),b_2(m),b_2(m+1)]$$

接收机如图 1-18 所示。

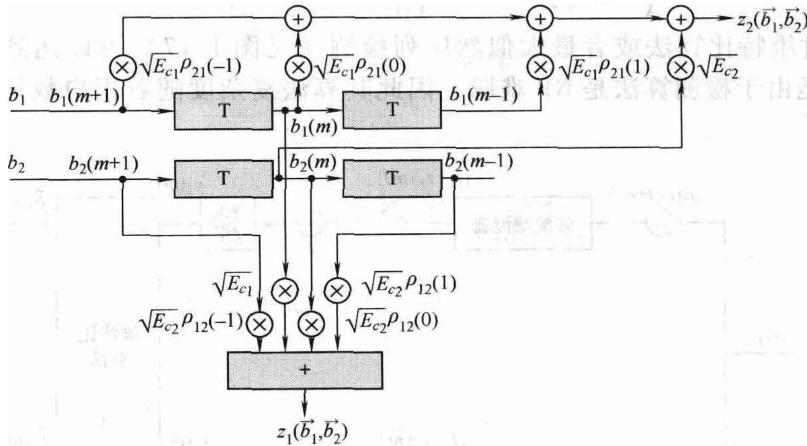


图 1-18 同步 CDMA 系统的抽头延迟线模型

注意：$\rho_{kh}(j)=\frac{1}{T}\int_{\tau_k}^{T+\tau_k}c_h(t-\tau_k)c_k(t+jT-\tau_c)\mathrm{d}t=\delta_J$

1.4.2 次优 DS/CDMA 接收机

由于最优多用户接收机复杂度随着用户数呈指数增长，在实际中很难实现，因此需要使用次优的接收机。

1.4.2.1 解相关接收机

两用户情况下，多用户接收机的输出可以很容易推广到更一般的 k 个用户系统

$$\boldsymbol{y}=\boldsymbol{RAB}+\boldsymbol{n}$$

式中，$\boldsymbol{R}$ 是互相关矩阵 $[\rho_{ij}]$；

$$A=\begin{bmatrix}\sqrt{E_{c_1}} & & 0\\ & \ddots & \\ 0 & & \sqrt{E_{c_k}}\end{bmatrix};$$

$$b=\begin{bmatrix}b_1(-M) & \cdots & b_1(M)\\ \vdots & & \vdots\\ b_k(-M) & \cdots & b_k(M)\end{bmatrix}$$

那么

$$R^{-1}y=Ab+R^{-1}n$$

忽略噪声项 t_0相当于对公式的右侧传输序列乘以比例。解相关接收机基于之前的线性变换，由以下步骤组成：

■ 匹配滤波器作为前端处理器产生 y；

■ 解相关计算 R^{-1}。

解相关多用户接收机的优点包括：

■ 低复杂度（在大多数情况下是线性复杂度）；

■ 不需要接收的幅度，因此能避免远近问题。

1.4.2.2　干扰消除（IC）

干扰消除是从接收的波形里串行消除干扰（见图 1-19）。通过干扰消除可以消除最强的检测信号，其步骤如下：

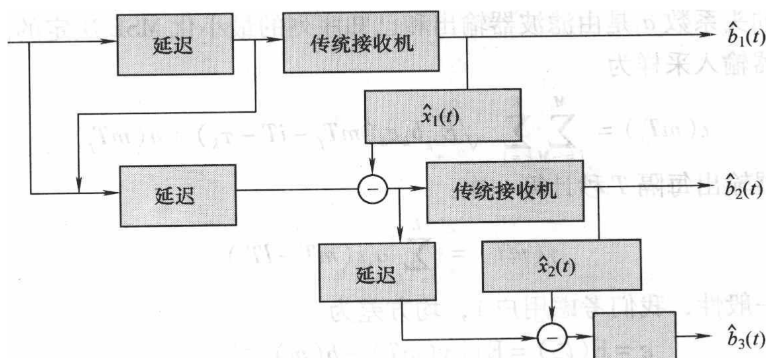


图 1-19　干扰消除

1）将用户信号排序 $\sqrt{E_{c_1}}>\sqrt{E_{c_2}}>\cdots>\sqrt{E_{c_k}}$；

2）使用传统接收机检测最强信号的用户；

3）重新产生信号最强用户的扩展频谱信号 $\hat{x}_k(t)=\sqrt{E_{c_k}}\,\hat{b}_k(t)c_k(t)$；

4）消除最强的干扰信号；

5）重复以上步骤直到消除所有的用户信号；

如果这种消除过程是串行进行的，我们称之为串行干扰消除（SIC），这个过程也可以是并行进行的，称之为并行干扰消除（PIC）。

1.4.2.3 自适应 MMSE 接收机

自适应最小均方差接收机不需要干扰信号的波形、定时和载波相位信息，只需要所要检测用户信号的定时和载波相位信息（见图 1-20）。

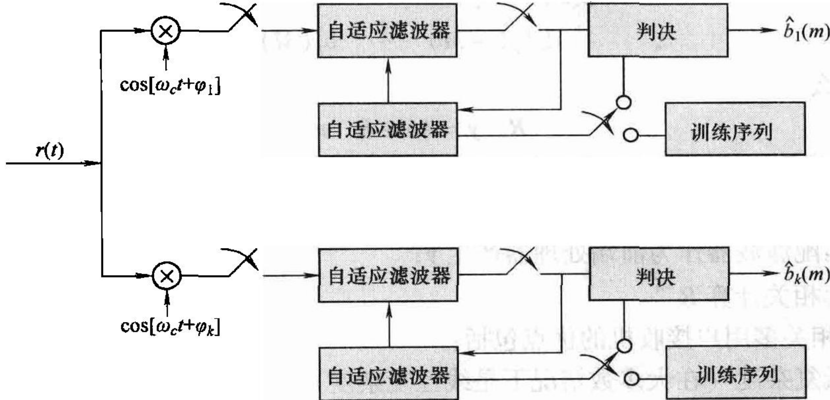


图 1-20 自适应 MMSE 接收机

通过分数滤波器可以极大地简化接收机的定时恢复，因此使得所要检测的用户信号对定时误差不敏感。自适应 FSE（分数间隔均衡器）滤波器由 $2L+1$ 个抽头组成。抽头系数 a_l是由滤波器输出和已知序列的最小化 MSE 决定的。

滤波器输入采样为

$$r(mT_f) = \sum_{i=-M}^{M}\sum_{k=1}^{K} \sqrt{E_{c_k}} b_k c_k(mT_f - iT - \tau_k) + n(mT_f)$$

滤波器输出每隔 T 秒计算一次：

$$y(mT) = \sum_{l=-L}^{L} a_l r(mT - lT_f)$$

不失一般性，我们考虑用户 1，均方差为

$$\varepsilon = \mathrm{E}(l_m^2) = \mathrm{E}\{[y(mT) - b(m)]^2\}$$

$$= \mathrm{E}\{[\boldsymbol{a}^{\mathrm{T}}\boldsymbol{r} - b(m)]^2\}$$

$$\boldsymbol{a} = (a_{-L}, \cdots, a_L)^{\mathrm{T}}$$

$$\boldsymbol{r} = [r(mT + LT_f), \cdots, r(mT), \cdots, r(mT - LT_f)]^{\mathrm{T}}$$

$$\varepsilon = \boldsymbol{a}^* \boldsymbol{D}\boldsymbol{a} - (\boldsymbol{a}^* \boldsymbol{f} + \boldsymbol{f}^* \boldsymbol{a}) - \sigma_b^2$$

式中，$\boldsymbol{D} = \mathrm{E}\ (\boldsymbol{r}\boldsymbol{r}^*)$；$\boldsymbol{f} = \mathrm{E}\{b^*(m)\boldsymbol{r}\}$；$\sigma_b^2 = \mathrm{E}[\,|b(m)|^2]$

单用户信道矢量$\boldsymbol{f}$ 的第 1 个元素f_e可以直接计算为

$$f_e = \mathrm{E}[b(m)r(mT+lT_s)] = \mathrm{E}[|b(m)|^2]c(lT_s)$$

最优的抽头系数为

$$\boldsymbol{a}_{\mathrm{opt}} = 1 - \boldsymbol{f}^* \boldsymbol{D}^{-1} \boldsymbol{f}$$

注意，在实际工程中，可能不能得到 $\boldsymbol{f}$ 和 $\boldsymbol{D}$，因此使得 LMMSE 更加实际。事实上，次优的多用户检测可以同样适用于 MIMO 接收机基本结构。回顾之前的

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n}$$

多用户检测问题也就等效于

$$\boldsymbol{y} = \boldsymbol{R}\boldsymbol{A}\boldsymbol{b} + \boldsymbol{n}$$

因此，如果我们把 $\boldsymbol{H}$ 和 $\boldsymbol{RA}$ 等效看待，次优多用户检测和 MIMO 接收机可以使用同样的接收机结构。

参考文献

[1] R.W. Chang, 'Synthesis of Band-Limited Orthogonal Signals for Multichannel Data Transmission', *Bell Systems Technical Journal*, **45**, 1960, 1775–1796.

[2] B.R. Saltzberg, 'Performance of an Efficient Parallel Data Transmission System', *IEEE Transactions on Communication*, **15**(6), 1967, 805–811.

[3] S.B. Weinstein, P.M. Ebert, 'Data Transmission of Frequency Division Multiplexing Using the Discrete Frequency Transform', *IEEE Transactions on Communication*, **19**(5), 1971, 623–634.

[4] A. Peled, A. Ruiz, 'Frequency Domain Data Transmission Using Reduced Computational Complexity Algorithms', *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'80), Denver, USA*, 1980, pp. 964–967.

[5] T. Pollet, M. Peeters, 'Synchronization with DMT Modulation', *IEEE Communications Magazine*, April 1999, 80–86.

[6] T. Pollet, P. Spruyt, M. Monenclaey, 'The BER Performance of OFDM Systems Using Non-Synchronized Sampling', *IEEE Global Telecommunications Conference*, 1994, pp. 253–257.

[7] J.M. Paez-Borrallo, 'Multicarrier Vs. Monocarrier Modulation Techniques: An Introduction to OFDM', *BWRC Retreat*, 2000.

[8] T. Pollet, M. Van Bladel, M. Monenclaey, 'BER Sensitivity of OFDM Systems to Carrier Frequency Offset and Wiener Phase Noise', *IEEE Transactions on Communication*, **43**(2), 1995, 191–193.

[9] J.G. Proakis, *Digital Communications*, 4th edition, McGraw-Hill, New York, 2000.

[10] T.A. Wilkinson, A.E. Jones, 'Minimization of the Peak-to-Mean Envelope Power Ratio of Multicarrier Transmission Schemes by Block Coding', *Proceedings of IEEE VTC '95*, July 1995, pp. 825–829.

[11] K.C. Chen, *Principles of Communications*, River, 2009.

[12] R. Van Nee, R. Prasad, *OFDM for Wireless Multimedia Communication*, Artech House Publishers, 2000.

[13] S.M. Alamouti, 'A Simple Transmit Diversity Technique for Wireless Communications', *IEEE Journal on Selected Areas in Communications*, **16**(8), 1998, 1451–1458.

[14] H. Jafarkhani, 'A Quasiorthogonal Space-Time Block Code', *IEEE Transactions on Communications*, **49**, 2001, 1–4.

[15] L. Zheng, D.N.C Tse, 'Diversity and Multiplexing: A Fundamental Tradeoff in Multiple-Antenna Channels', *IEEE Transactions on Information Theory*, **49**, 2003, 1073–1096.

[16] R.W. Heath Jr., A. Paulraj, 'Switching Between Spatial Multiplexing and Transmit Diversity Based on Constellation Distance', *Proceedings of Allerton Conference on Communication, Control and Computing*, October 2000.

[17] H. Skjevling, D. Gesbert, N. Christophersen, 'Combining Space Time Block Codes and Multiplexing in Correlated MIMO Channels: An Antenna Assignment Strategy', *Proceedings of Nordic Signal Processing Conference (NORSIG)*, June 2003.

[18] A. Gorokhov, D. Gore, A. Paulraj, 'Diversity Versus Multiplexing in MIMO System with Antenna Selection', *Allerton Conference*, October 2003.

[19] V. Tarokh, N. Seshadri, A. Calderbank, 'Space-Time Codes for High Data Rate Wireless Communications: Performance Criterion and Code Construction', *IEEE Transactions on Information Theory*, **44**, 1998, 744–765.

[20] Y. Xin, Z. Wang, G.B. Giannakis, 'Space-Time Diversity Systems Based on Linear Constellation Precoding', *IEEE Transactions on Communications*, **49**, 2001, 1–4.
[21] X. Ma, G.B. Giannakis, 'Complex Field Coded MIMO Systems: Performance, Rate, and Trade-offs', *Wireless Communication and Mobile Computing*, 2002, 693–717.
[22] T.J. Jung, K. Cheun, 'Design of Concatenated Space-Time Block Codes Using Signal Space Diversity and the Alamouti Scheme', *IEEE Communication Letters*, **7**, 2003, 329–331.
[23] S. Verdu, *Multiuser Detection*, Cambridge University Press, 1998.

第2章　软件无线电

2.1　软件无线电架构

目前已经有很多不同的无线通信系统广泛存在并使用在不同的应用场景。图2-1列出了一些熟知的国际无线通信标准，包括无线体域网、无线个域网、无线局域网和无线城域网、无线广域网，这些无线通信系统有着不同的应用场景和系统参数。事实上，对于蜂窝系统，目前使用的也有很多不同的系统，例如合法的GSM、GPRS和EDGE、3GPP宽带码分多址（WCDMA）及其升级版本HSDPA、HSUPA，以及未来的3GPP长期演进（LTE），上述也只是考虑了空中接口技术。图2-2描述了可能的蜂窝系统空中接口的融合，我们可以很容易地看到已经使用或即将使用的这些系统，包括时分多址接入（TDMA）、码分多址接入（CDMA）和正交频分复用接入（OFDMA）。这些空中接口技术可能在不同的区域内共存，也可能是同时在同一个位置共存。因此需要设计一个灵活的终端设备允许用户来合适地使用无线通信。

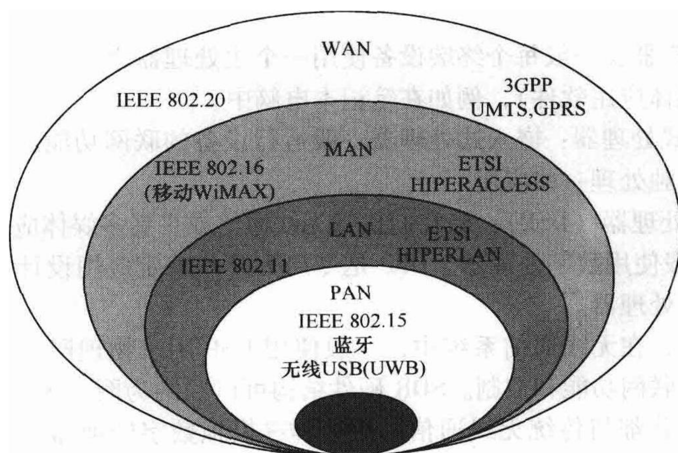


图2-1　全球无线通信标准

由于在通信发展的早期阶段，典型的无线通信系统是通过某些专用硬件来实现的，例如根据特定的通信系统参数设计使用的专用集成电路（ASIC），这种方式和我们现在用计算机在同样的计算设备上执行不同的应用程序很不相同。软件

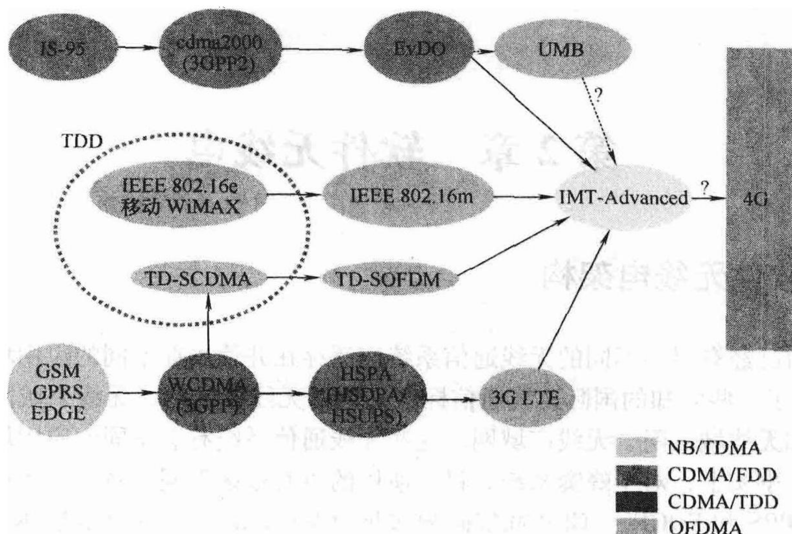


图 2-2 蜂窝系统空中接口技术的演进

无线电（SDR）能在同一个硬件平台通过软件来实现多个无线通信系统，因此在之前描述的场景（多个系统在同一个位置共存或者全球共存）下具有广阔的前景。

对目前的终端设备，一般使用的处理器根据特定的应用和无线通信/网络分为以下三类：

1）微处理器：一般每个终端设备使用一个主处理器用于基础应用（包括编程或执行多媒体应用软件），例如在笔记本电脑中。

2）嵌入式处理器：嵌入式处理器一般运行设备的联网功能，有时也和主处理器一起或单独处理一些应用。

3）数字处理器（DSP）：当我们执行无线通信或扩展多媒体应用时，例如音频/视频，一般使用数字处理器。DSP 是专门为某些特定应用设计的具有强大处理功能的专用处理器。

换句话说，在无线通信系统中，一般使用 DSP 用于物理层基带处理，嵌入式处理器用于联网功能和控制。SDR 硬件结构可以归纳为图 2-3。天线、功率放大器和射频电路都与传统无线通信硬件一样，模拟数字转换器（ADC）把接收机的模拟波形转换为数字采样，DAC 则把发射机的数字采样变成模拟波形。SDR 的灵活性主要体现在基带部分的可编程性。由于 SDR 一般工作在多个频带上，我们可以使用一组专门的射频，也可以使用可调谐的宽带射频，从而可以工作在多个频带。

在实现物理层通信接发机（发射机和接收机）时，射频部分和功率放大器

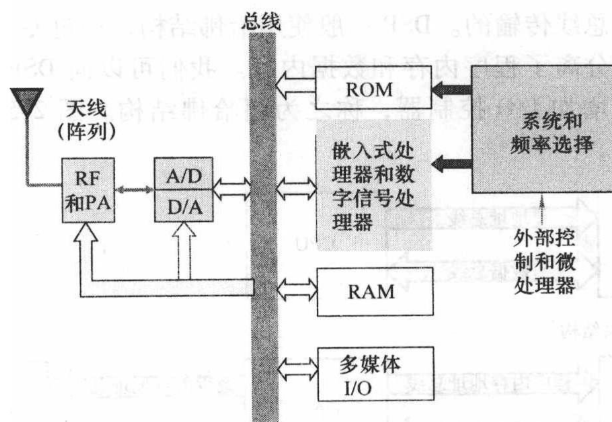


图2-3 SDR设备的硬件结构

都是由模拟电路来实现的，基带部分由DSP实现，其中最主要的设计是把每个基带通信模块变换成DSP可编程算法。高层网络功能一般是由嵌入式处理器程序来实现的。设备（主）微处理器和应用处理器则负责多媒体和用户接口功能（见图2-4）。



图2-4 SDR设备的软件结构

2.2 数字处理器和SDR基带结构

与典型的通用处理器处理数据不同，数字处理器将数字计算使用于特定的应用，例如控制、媒体处理、通信等，相应的运行速率更高。目前很多无线通信产品都是用DSP实现的。

典型的通用处理器使用熟知的冯诺依曼体系结构，内存由指令共享，数据

是通过单个数据总线传输的。DSP 一般使用哈佛结构，通过专用程序内存总线和数据内存总线分离了程序内存和数据内存。我们可以向 DSP 的 CPU 中增加指令缓存，并且增加 I/O 控制器，称之为超哈佛结构。图 2-5 表示了这三种结构。

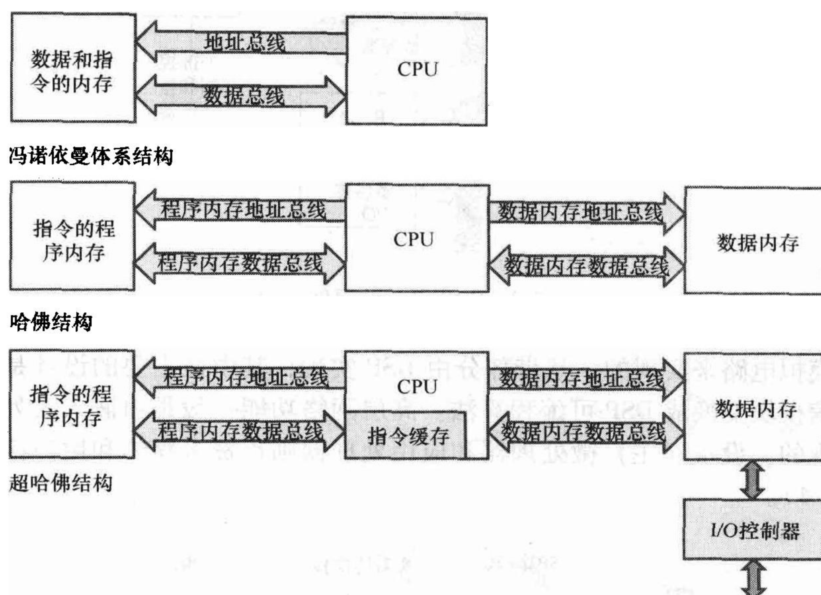


图 2-5 处理器结构

由于多媒体或通信的信号波形一般是模拟信号，需要通过 A/D 转换把模拟波形编码成数字采样，从而方便 DSP 的处理。数字信号处理使用 DSP 有两种方式：浮点和定点，主要区别在于存储和操作的数据格式。定点 DSP 中每个采样数据是由固定比特数表示的，一般是 16bit、32bit 或 64bit，有些 DSP 是用 24bit。对于 16bit 定点 DSP，每个采样可能有 $2^{16}=65536$ 个无符号整数值。浮点 DSP 一般使用 32bit 甚至 64bit 来表示每个采样。在一般的 32bit 浮点格式中，能表示的最大数和最小数分别为 $\pm3.4\times10^{38}$ 和 $\pm1.3\times10^{-38}$。

DSP 使用汇编语言编程，特别是我们想优化 DSP 性能时（也就是加快运行速度时）。目前的 DSP 一般支持上百条执行指令。然而，一般的 DSP 开发系统支持更高级的 C 语言，相比汇编语言，C 语言更加方便编程，但是 DSP 效率更低。定点 DSP 处理能力一般是用每秒百万指令（MIPS）来衡量，而浮点 DSP 的处理能力是用每秒百万次浮点操作（MFLOPS）来衡量。目前的 DSP 可以支持几百 MFLOPS，比通用处理器的处理能力要强得多。为了进一步达到数字信号处理对处理能力和多任务的需求，还可以在一个 DSP 中包括多个算术逻辑单元

（ALU），并且在一个系统中集合多个DSP，称之为多处理或并行处理。图2-6描述了并行DSP的通用系统架构。总线在DSP和内存之间交互数据中起着重要的作用，仲裁器可以共享内存，也可能控制内存。

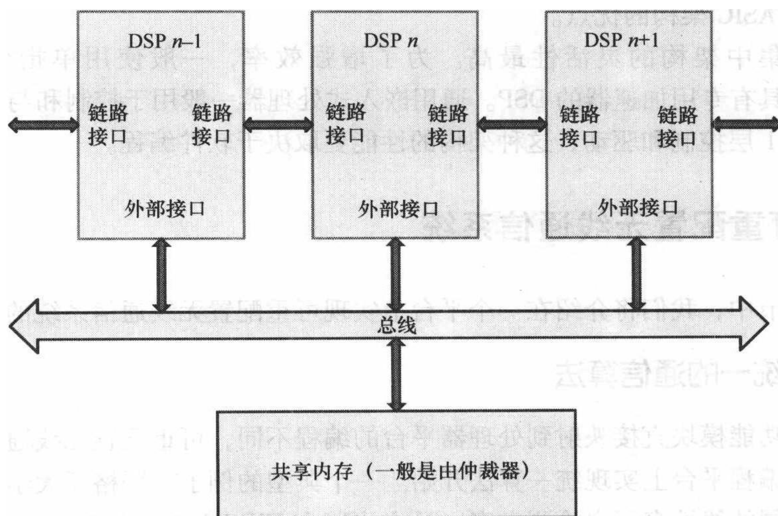


图2-6 多处理或DSP架构

为了用DSP执行软件无线电的基带部分，可以考虑如图2-7所示的几种系统架构以平衡灵活性、功耗、软件复杂度和成本（芯片大小、系统集成等）。

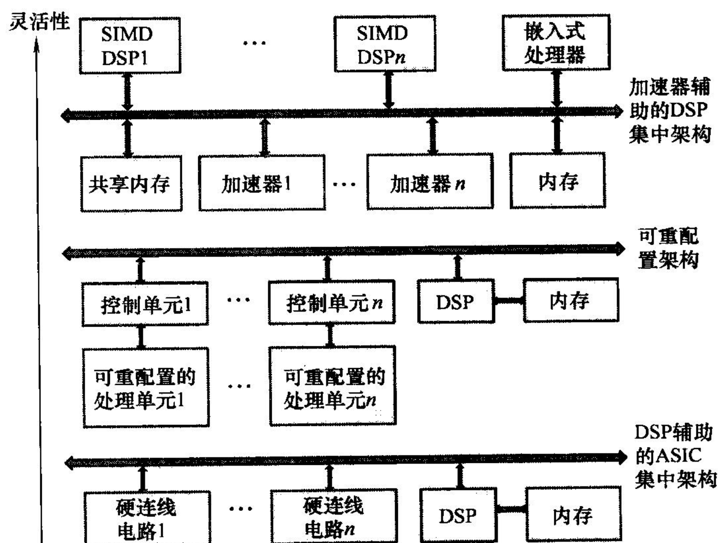


图2-7 SDR基带硬件系统架构

ASIC 集中架构是实现多标准设备经典的硬件架构，其中可能有 DSP（或嵌入式处理器）辅助 ASIC 芯片处理。软件复杂度完全在 1 层控制或驱动。

可重配置架构比 ASIC 方案更加灵活，通过控制合适设计的数据处理单元以保持相对 ASIC 架构的优点。

DSP 集中架构的灵活性最高。为了增强效率，一般使用单指令多数据（SIMD）具有专用加速器的 DSP。通用嵌入式处理器一般用于控制和与上层的接口。除了 1 层控制和驱动，这种架构的性能更取决于软件编程。

2.3 可重配置无线通信系统

在本节中，我们将介绍在一个平台上实现可重配置无线通信系统的方法。

2.3.1 统一的通信算法

与把功能模块直接映射到处理器平台的编程不同，可重配置无线通信系统从在一个可编程平台上实现统一算法开始。一个典型的例子是罗格斯大学的研究者执行的不同的线性多用户检测方案（论文出版在 IEEE JSAC 1999）[2]。正如在之前章节（1.4 节）中指出的，我们有很多次优的多用户检测接收机结构，在这里只考虑以下三种熟知的线性多用户检测接收机：

1）匹配滤波器（ME）接收机，一种多用户检测接收机的简化版传统接收机；

2）解相关（DC）接收机；

3）最小均方差（MMSE）接收机。

我们考虑同步系统的 K 个用户，第 k 个用户的匹配滤波器输出可以表示为

$$y_k = A_k b_k + \sum_{j \neq k} A_j b_j \rho_{jk} + n_k \tag{2-1}$$

式中，A_k是第 k 个用户的幅值，$b_k \in \{-1, 1\}$ 表示第 k 个用户的数据比特；$n_k = \sigma \int_0^T n(t) s_k(t) \mathrm{d}t$，$s_k(t)$ 是第 k 个用户归一化能量的信号波形，ρ_{jk}表示信号波形 $s_j(t)$ 和 $s_k(t)$ 的互相关，是协方差矩阵 $\boldsymbol{R}$ 的元素。系统中用户的匹配滤波器输出可以用矢量形式表示为

$$\boldsymbol{y} = \boldsymbol{RAb} + \boldsymbol{n} \tag{2-2}$$

式中，$\boldsymbol{y} = [y_1, \cdots, y_K]^{\mathrm{T}}$，$\boldsymbol{b} = [b_1, \cdots, b_K]^{\mathrm{T}}$，$\boldsymbol{A} = \mathrm{diag}\{A_1, \cdots, A_K\}$，$\boldsymbol{n}$ 是零均值高斯随机矢量，其协方差矩阵为 $\sigma^2 \boldsymbol{R}$。线性多用户检测方案可以通过对接收矢量 $\boldsymbol{y}$ 的线性变换来实现。解调过程可以简化为

$$\hat{b}_k = \mathrm{sgn}(\boldsymbol{Ly})_k$$

式中，L 是对应的线性变换。对线性多用户检测接收机，我们考虑

$$L_{\mathrm{MF}} = 1$$

$$L_{\mathrm{DC}} = \boldsymbol{R}^{-1}$$

$$L_{\mathrm{MMSE}} = [\boldsymbol{R} + \sigma^2(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}]^{-1}$$

众所周知，这三种接收机的误码率性能是 $BER_{\mathrm{MMSE}} \leqslant BER_{\mathrm{DC}} \leqslant BER_{\mathrm{MF}}$。图 2-8 归纳了这三种接收机处理流程的逻辑步骤，这些计算可以根据通信理论重用在多系统（也就是多个多用户检测接收机）的配置和算法上。

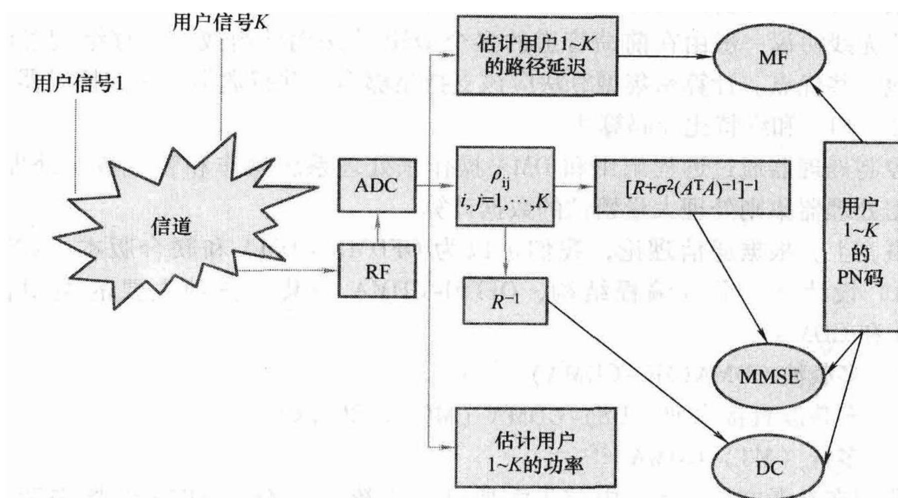


图 2-8　可重配置线性多用户检测的逻辑功能

2.3.2　可重配置 OFDM 实现

可以通过以下方式实现可重配置通信系统：①可重配置元器件，例如可调谐射频电路和可编程门阵列（FPGA）；②不同类型元器件的组成，例如处理器、ASIC 或硬连线电路、内存和模拟设备等。第一种方法一般成本较高，包括功耗、控制负荷或设备尺寸。我们需要使用第二种方法的元器件构建一个可重配置通信系统。华盛顿大学开发的 RaPiD 可重配置结构就是通过提供可重配置管道数据通道和可重配置控制逻辑而进行通信和信号传送，这是结合了可编程 DSP 和 ASIC 很好的实例。可以在这种可重配置架构的基础上实现一个 OFDM 系统作为示例。可重配置数据通道有一些功能模块，包括 ALU 和乘法器，当然还包括很多寄存器，用于存储每个循环时功能模块所需要的信息，此外还有一些嵌入式（随机接入）内存，用于存储在计算中需要重复使用的数据。

2.3.3 可重配置 OFDM 和 CDMA

可重配置系统架构的下一个阶段是在一个平台上演示多个无线系统。Lin 等在参考文献［4］提出了由一个控制器和一些超宽 SIMD 处理单元（PE）组成的 SDR 硬件平台，数据通信通过直接存储器访问（DMA）指令实现。这种系统结构一般用于高端通用计算，例如 IBM 的 Cell 处理器。WCDMA（3G 蜂窝网）和 IEEE 802.11a（WLAN）都通过这种可编程架构实现用于演示的双模系统。实现无线系统首先要分析无线协议，总共有四种主要功能：滤波、调制、信道估计和纠错。无线协议一般由在前馈管道的多个 DSP 算法内核组成。缓存结构可以带来其他一些优点。计算密集型算法应该支持足够多的并行数据，例如搜寻器、低通滤波、FFT 和维特比译码算法。

控制处理器通过远程调用和 DMA 操作来处理系统操作和管理数据处理器，而数据处理器集则处理大量繁忙的数据计算。

事实上，根据通信理论，我们可以为 OFDM、CDMA 和联合版本（OFDM-CDMA）设计出一个可编程结构，OFDM-CDMA 一共有三种主要的类型合并 OFDM 和 CDMA：

1）多载波 CDMA(MC-CDMA)[9,10]；

2）多载波直接序列（DS）-CDMA（MC-DSCDMA）[11]；

3）多音（MT）-CDMA[12]。

我们在参考文献［5］中首先需要开发出统一的 OF-CDMA 结构来调节参数，从而实现以上三种系统中的任意一种，可以进一步简化为实现简单的 OFDM 和 CDMA。然后，我们可以调节参数以实现可编程 OFDM-CDMA 平台，从而构成现代无线通信系统的基础。

2.4 数字无线电处理

现代无线通信系统的基站和移动台都是通过片上系统（SoC）技术实现的。由于在 20 世纪 90 年代的巨大突破，射频部分都可以用 CMOS 芯片来实现，整个发射接收机可以嵌入到单个片上系统实现。根据著名的摩尔定律，集成电路的门电路密度每隔 18 个月增长一倍。换句话说，对于给定的无线通信系统，SoC 芯片大小（一般与成本和模块大小成比例）可以每隔 18 个月降低一半。不幸的是，这种结论只对数字部分成立，而对模拟部分不适用。在几年内，无线通信的片上系统上主要部分可能是射频部分的模拟电路。一个开创性的想法是把传统的射频电路用数字处理器取代，称之为数字无线电处理器（DRP）。最近德州仪器（TI）已经把这个想法变成商业产品[15-17]。

2.4.1 传统射频

超外差（HF）和直接变换（DC）是两种最常见的射频系统结构。这两种下变频模型很容易分析和理解，但是需要更多外部元件，从而导致复杂度更高，因此效率较低并不适合于片上系统。另一方面，直接变换结构相对比较简单，因此更适用于片上系统，但是实际实现上存在一些挑战。

2.4.1.1 接收机

图2-9从实际的角度描述了一般的射频系统。在图2-9b中，接收到的射频信号通过低噪放大器（LNA）放大后直接变换到基带的同相I和正交Q信号。通过芯片低通滤波器和基带的可变增益放大器（VGA）实现信道选择和增益控制。超外差接收机的信道选择时，通过把射频信号下变频到混合中频，并通过中频表面声波（SAW）（或者晶体滤波器）滤波器。在多个频带/标准工作的接收机具有不同的工作频率，因此需要不同的中频滤波器，即使是在小型的手机上。因此，直接变频接收机结构具有一些优势。

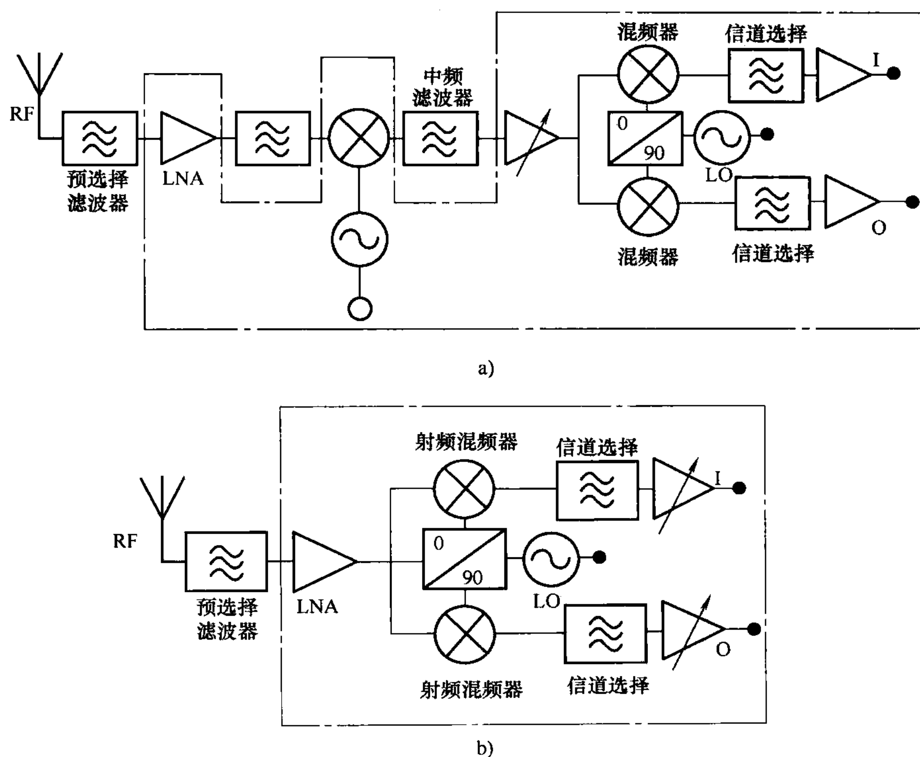


图2-9　a）超外差和b）直接变频结构

在基带进行信道滤波不仅可以使得芯片内实现高效节能的滤波成为可能，而且也不再需要外部被动式滤波器，从而减少了电路板大小和成本（见图2-9b）。

设置直接变频接收机（DCR）的中频为0，在超外差结构的情况下，没有镜像频率，因此DCR中LNA和混频器之间不再需要级间镜像抑制滤波器。

不再需要传统超外差接收机架构中使用的甚高频电压控制振荡器（VHF VCO）和锁相环（PLL），从而减少了芯片大小和成本，此外消除了VHF VCO带来的相位噪声问题。

然而，DCR结构中还有一些关键的问题，例如直流偏置、二阶线性、本振泄露、增益和相位不平衡等问题存在。

2.4.1.2 **发射机**

正如其名字所示的，直接变频发射机是把基带IQ信号直接调制到要发射的频率的射频载波上。如图2-10b所示，通过射频自动增益控制（AGC）实现可变的射频增益，调制后的射频信号放大到所需的功率输出。射频SAW滤波器可以用于在信号经过功率放大器之前抑制接收机接收频带的本底噪声。这与传统的两次变频发射机不同，因为传统的两次变频使用了两个本振源，通过中频和射频可变增益放大器实现所需的动态范围。

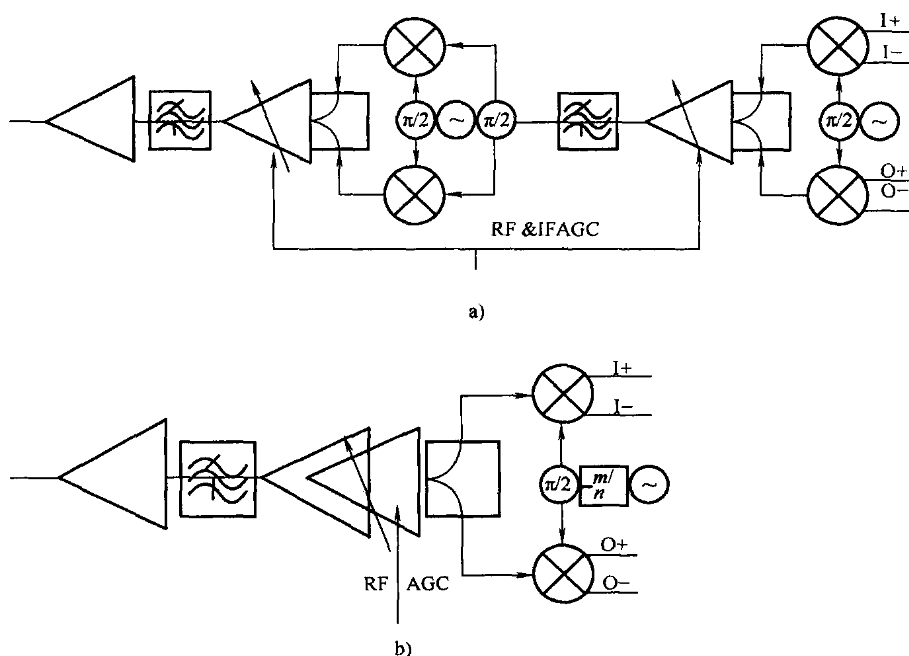


图2-10 a）超外差和b）DCT结构

目前的 CDMA 终端中的发射机架构是根据图 2-10a 所示的双变频传输链，其中基带 IQ 信道调至到中频载波之后，通过一个动态范围大于 80dB 的可变增益放大器。放大后的中频信号通过简单的 LC 调谐元件进行滤波，然后通过镜像抑制混频器和可变增益驱动模块进行上变频。增益划分用于功率控制、降低噪声等级与线性优化功率效率。商用的射频芯片一般集成了 2 ~ 3 个锁相环和 1 个甚高频压控振荡器。传输 IQ 信号的信道滤波和波形成形一般都在基带进行处理。CDMA/WCDMA 发射机的中频滤波只是用于抑制在中频可变增益放大阶段产生的接收频带噪声。中频滤波一般是用调谐到发射中频上的简单 LC 并行谐振电路来实现。这种滤波器提供的频率滚降使得发射中频 + 双工频率间隔（WCDMA 的双工频率间隔为 190MHz，而 USPCS CDMA 的双工频率间隔为 80MHz）上的接收频带噪声功率衰减到足够低的水平。

DCT 与双变频发射链相比有以下一些优点：

1）DCT 不需要使用片上中频 VCO + PLL + 调谐电路元件，以及用于减少接收频带噪声的外部 LC 中频滤波器，因此减少了元件数量。

2）DCT 由于规划设计可以用于具有不同双工间隔和频率范围的多模式系统（例如 GSM/WCDMA）。而对双变频发射机来说很难实现。片上调谐和可编程性使得不同的移动标准的功能模块之间有可能实现复用，从而减少了芯片大小和成本。

3）由 $m \cdot \mathrm{IF} \sim n \cdot \mathrm{LO}$（IF 为中频，LO 为本振泄露频率）造成的杂散辐射可以完全消除，因为中频为 0。

然而，DCT 仍然有以下一些问题：

1）IQ 相位和增益失衡；

2）带内本底噪声；

3）压控振荡器频率出现偏移；

4）动态范围；

5）功耗。

在基带（模拟/数字）和射频部分分配合适的增益可以获得动态范围和相应的性能折中，例如数字部分的可变增益（需要较高的 DAC 分辨率）、模拟基带部分的可变增益（导致混频器的本振泄露增大）和射频部分的可变增益（导致高的功耗）。

2.4.1.3　频率合成器

射频频率规划中最重要的组成部分是根据锁相环（PLL）的本地振荡（LO），它产生所需的载频，有时控制整个发射接收模块。图 2-11 说明了最常见的商业产品中的 PLL。

相位/频率检测器（PFD）根据测量的相位边缘时间差来估计频率参考 f_{ref} 输

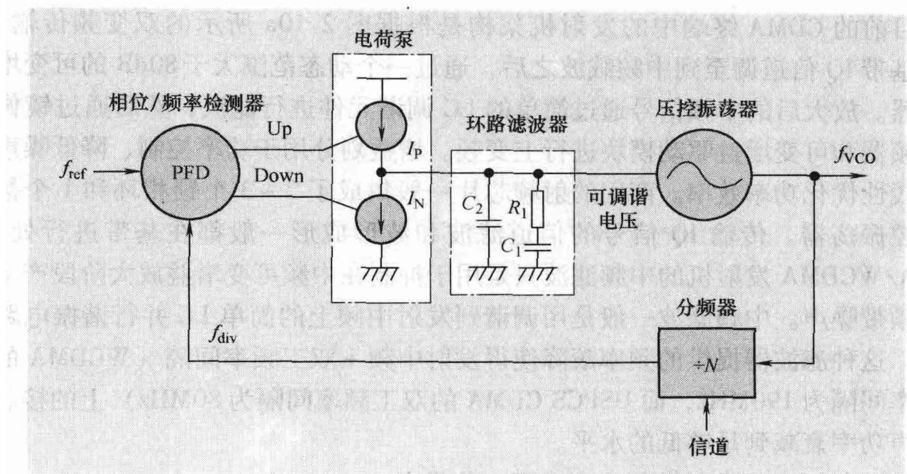


图 2-11　一种典型的基于电荷泵 PLL 的本地振荡

入和 N 分压控振荡器（VCO）时钟 f_{div} 之间的相位差。PFD 产生一个 Up 或者 Down 脉冲，该脉冲的宽度与测量的时间差成正比。该信号反过来又在电荷泵模块产生一个占空比成比例的脉冲电流 I_P 或 I_N。在闭环滤波器中，这个电流转换成 VCO 的调谐电压。然而，电荷泵锁相环有以下缺点：

1）由于电荷泵在每次相位比较时会产生周期短时脉冲，该周期脉冲会调制 VCO 输出频率，因此造成频率杂散，从而影响发射机和接收机性能。

2）通过减少锁相环的带宽可以降低频率杂散，但是也会降低瞬态响应，因此需要在两者间折中。

3）电荷泵锁相环是模拟的，因此不适合集成到微米级的 CMOS 技术。

4）电荷泵的电流源需要精确匹配，由于相对低的晶体管电阻和特别薄的栅氧化层，用于锁相环滤波器的大集成电容很难集成到微米级 CMOS。

5）VCO 的电压-频率特性中逐渐增加的非线性使得 PLL 带宽不稳定。

2.4.1.4　A/D 转换器

射频动态范围中另一个最重要的部分是 A/D 转换器，它主要是把射频输入信号在出现很大干扰时从模拟信号变换到数字信号。通用的商用无线标准所需的动态范围要远超过单个 ADC 所能实现的动态范围。而临近信道的干扰功率（其中不包含有用的信息）远大于要检测的信号功率，检测器所需的信噪比只有 8～10dB。以 GSM 系统为例，有用信号 3MHz 偏移的干扰是 -23dBm，而灵敏度是 -102dBm。虽然具有低噪放大器（LNA）和混频器的 GSM 无线电接收机可以通过 200kHz 频带内动态范围为 90dB 的 ADC 提供大约 20～30dB 的增益，但是在实现上这样不现实，并且也不是必须这样实现。因为实际上在超过 80dB 动态范围

时，收益递减定律开始发挥作用，每增加 1dB 的动态范围并不与所能改善的 ADC 性能成比例（尤其是在多标准无线电系统中）。因此，一般的接收机设计时都在滤波器阶段后增加一系列增益阶段以降低对 ADC 性能的需求。增益阶段减少了在之后阶段中噪声的影响，而滤波器阶段抑制了干扰/阻塞的额外信息，因此为接收机之后的处理阶段提供了更多增益，而不会让接收机饱和失真。

直流偏移减少了 ADC 可达到的动态范围，因为偏移导致接收信号波形出现失真，从而导致非线性。一个或多个直流偏移矫正（DCOC）电路可以在接收机中用于消除接收链路中的直流偏移，这对直接变换接收机尤为重要。在移除临近信道的干扰后，消除残留偏移，信号通过检测电路。此外可能还需要信号前端处理，例如让信号在硬判决之前通过匹配滤波器和均衡器。

2.4.2 基于数字无线电处理（DRP）的系统架构

考虑到单片机融合的趋势和先进的数字 CMOS 处理技术（深亚微米级 CMOS 处理），一类新的面向模拟和射频设计者的深亚微米级 COMS 和单片机电路可以定义如下：

在深亚微米级 CMOS 处理中，数字信号边缘过渡的时域分辨率比模拟信号的电压分辨率高。

在这种环境下成功的设计方法应该强调以下几点：

1）MOS 晶体管的快速切换特征或者很高的 f_T（在该处理中分别为 40/s 和 100GHz）；定时转换高速时钟和精确控制；

2）目前数字逻辑的高密度工艺水平使得数字功能模块十分便宜；

3）精确的光刻技术使得小的设备尺寸和精确的设备匹配成为可能。

然而，在设计时还必须避免以下问题：

1）模拟设计时通常使用的偏置电流；

2）依赖于电压分辨率；

3）内存和数字电路不需要的非标准设备。

利用先进数字信号处理技术快速和精确的优点，德州仪器（TI）首先提出了几乎全数字化的单片机技术，称为数字无线电处理（DRP），DRP 把射频处理从模拟域变到数字域，因此射频处理可以实现高度集成。其中包含以下改进：

1）采样数据处理技术，一般用于直接射频采样，用时间离散模拟域处理取代传统的时间连续模拟域混频器的相乘，例如快速下变频和避免使用能耗较大的传统主动混频器。

2）交换电容滤波器，数字信号处理概念在滤波器中用于抑制噪声和临近

信道干扰，从而实现高选择性的射频滤波器，并且避免使用传统的外置滤波器。

3）过采样转换器，过采样技术不仅仅降低了 A/D 转换分辨率，还通过噪声整形的特点降低了对滤波器的抗混叠的限制。

4）全数字锁相环（ADPLL），基于数控振荡器（DCO）和时间数字转换器（TDC），因此避免使用传统的压控振荡器模拟调谐，并提供了更精确的数字相位检测器取代传统电荷泵锁相环中的模拟相位频率检测器。

2.4.2.1　接收机

在图 2-12 中，接收的射频信号通过低噪放大器（LNA）放大，分成 I/Q 支流，然后通过互导放大器（TA）变换成电流。电流下变频到可编程低中频（默认为 100kHz），再通过工作在本振频率的采样电容。考虑到正负频率，输入信号为射频载波的奈奎斯特频率采样。在通过 sinc 滤波响应初始抽取之后，射频采样通过一系列无限脉冲响应（IIR）滤波器，从而抑制相邻信道的干扰。这些信号处理过程是通过多抽头直接采样混频器（MTDSM）实现的。Σ-ΔA/D 转换器包含有前端增益，把模拟信号转换成数字信号形式。反馈控制单元（FCU）向 MTDSM 提供了单比特反馈，从而确定 MTDSM 的通用模式电压，同时抵消了差分偏移。I/Q ADC 的输出通过数字接收（DRX）链，进行下变频到零中频，抑制邻近信道干扰等操作。因此，一方面，这是低中频架构类似于 DCR 结构，具有和 DCR 同样实现上的问题。另一方面，这种 DRP 系统和传统的模拟射频系统之间存在着折中。

使用直接射频采样技术来下变频输入的射频信号为数字-模拟概念带来了更多快速运算，因此消除了传统混频器的问题（例如，镜像抑制问题或互调产物）。但是之后还需要滤波，以抑制或消除这种采样接收机中的噪声累积和噪声抖动问题。

使用中频数字化技术把模拟信号变换到数字域，让 ADC 更靠近天线端，并把大多数功能模块放到数字域来处理（例如信道选择滤波、零中频下变频、纠错和解调），因此消除了由于模拟实现带来的灵敏度问题，例如设备匹配、相位噪声、环境灵敏度和随之变化的性能。但是，在大的干扰和阻塞出现时，高速 ADC 的分辨率和动态范围会下降，功耗随着采样速率增加而增大，甚至影响系统中的其他设备。

虽然由于直接射频采样和中频数字化，避免了使用中频滤波器和镜像抑制滤波器，但是仍需要一些抽取滤波器来降低数据速率，如图 2-12b 中。

此外，数字域需要包含有外部反馈回路，如图 2-12a 中的反馈控制单元（FCU）和 1 比特 DAC，从而在不同的模式下工作，以抵消直流偏移、消除干扰和阻塞、改进线性度。

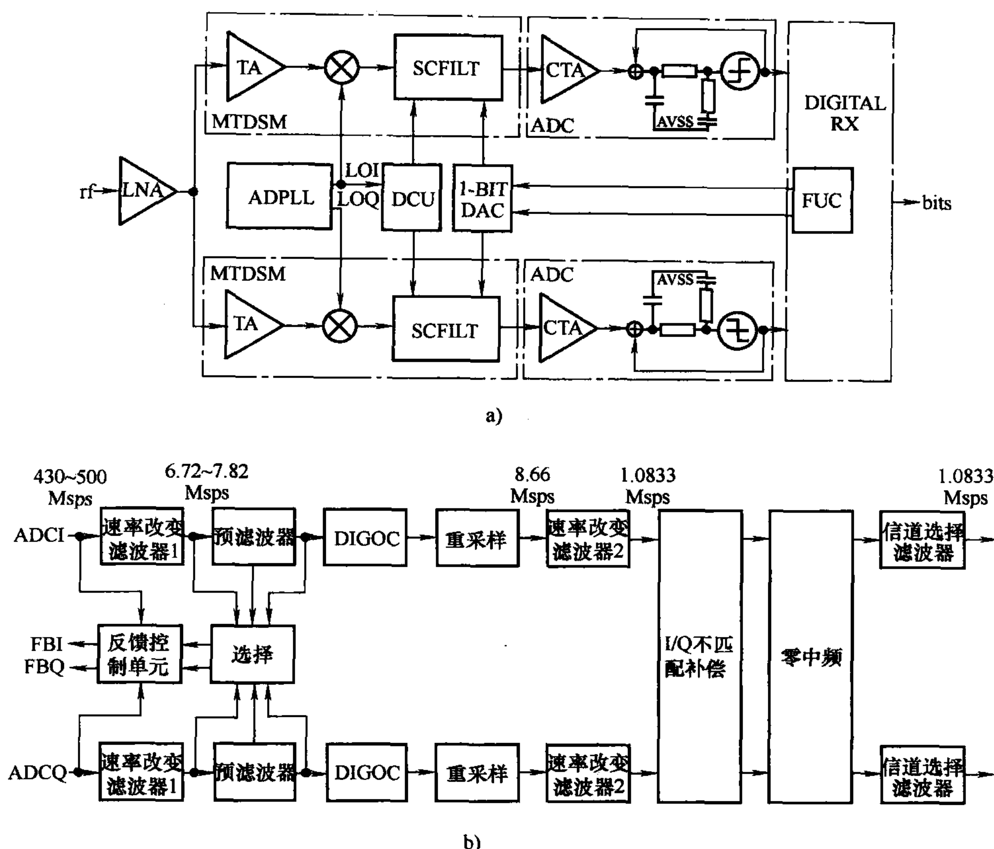


图2-12 a）DRP接收机和b）数字接收链路的简化框图

2.4.2.2 发射机

图2-13所示的是一个适合用深亚微米级CMOS技术实现的发射机。它在极坐标域进行正交调制，还为接收机产生本地振荡。系统中的所有时钟都直接从发射机中获得。这种结构利用了先进CMOS技术的高速、高密度的特点，并且避免了与电压余量有关的问题。ADPLL取代了传统的射频合成器结构，基于压控振荡器和相位/频率检测器（PFD），电荷泵结合了数控振荡器（DCO）和时间数字转换器（TDC）。所有的输入和输出信号都是数GHz频率的数字信号——40ps的上升时间几乎是完好的方形波。射频频率的完全数字化控制使得锁相环也可以实现数字化。

全数字化锁相环的中心是数控振荡器。数控振荡器工作在1.6～2GHz频率的两倍频率。DCO调谐电容划分为很多数字选择的小电容。先进的光刻技术可以生产精确的可变电容（变容器）——大约每步40aF电容，等于控制让250个电子进入或离开LC谐振回路。虽然每步控制的电容很小，在2GHz射频输出时

a)

b)

图 2-13 DRP 极性发射机

a）幅度调制路径极性发射机 b）基于全数字 PLL 的极性发射机

的频率步进为 10 ~ 20kHz，对于无线应用来说这个频率步进过大。因此，晶体管快速切换的能力是通过可编程控制高速（225 ~ 900MHz）振动的 250 个电子进入精确的电容器来实现的。高/低电容的占空比决定了平均谐振频率分辨率不低于 1kHz。所有的变容器都是用工作在 *C*- *V* 曲线的平坦区的 N-Poly/N-Well MOSCAP 设备实现的。

2.4.2.3 全数字化锁相环

图 2-13b 所示的 ADPLL 工作在数字同步定点相位域。可变相位 R_v ［i］ 是通

过统计 DCO 振荡时钟 CKV 的时钟上升边缘来确定的：

$$R_v[i] = \sum_{l=0}^{i} 1$$

i 表示 DCO 的时钟边缘。参考频率 FREF 采样可变相位 $R_v[k]$，其中 k 是参考频率边缘下标，$R_v[k]$ 是与归一化 TDC 输出相关的定点值。TDC 测量并量化 FREF 和 DCO 边缘的时间差。数字频率检测器从 FCW 中提取采样差分可变相位。频率误差采样 $f_E[k]$

$$f_E[k] = FCW - [(R_v[k] - \varepsilon[k]) - (R_v[k-1] - \varepsilon[k-1])] \tag{2-3}$$

累积，从而得到相位误差采样 $\phi_E\ [k]$

$$\phi_E[k] = \sum_{l=0}^{k} f_E[k]$$

相位误差通过四阶 IIR 滤波器，并且通过比例环路衰减 α。并行系数 ρ 增加了融合项，从而形成Ⅱ型环路特性，进而抑制 DCO 闪烁噪声。

IIR 滤波器是由 4 个单阶滤波器级联而成，每个滤波器都满足式（2-4）：

$$y[k] = (1-\lambda)y[k-1] + \lambda x[k] \tag{2-4}$$

式中，$x[k]$ 是输入，$y[k]$ 是输出，λ 是可配置系数。四级 IIR 滤波器衰减了参考信号，TDC 量化噪声的斜率为 80dB/dec，主要是为了满足 400kHz 偏移时的 GSM 频率功率需求。通过滤波器和衰减的相位误差采样乘以 DCO 增益归一化因子 $f_R/\hat{K}_{\mathrm{DCO}}$，其中 f_R 是参考频率，$\hat{K}_{\mathrm{DCO}}$ 是估计的 DCO 增益，环路特征和调制与 $\hat{K}_{\mathrm{DCO}}$ 独立无关。调制数据输入 ADPLL 的两个地方，用于直接频率调制。动态环路带宽控制使用无缝换挡机制来减少设置时间。它在锁定频率过程中改变了环路衰减 α 倍，并且相位误差增加了直流偏置（$\alpha_1/\alpha_2 - 1$）ϕ_1，其中下标 1 和 2 分别表示事件前和事件后。当然，由于相位是连续的，$\phi_1 = \phi_2$。FREF 输入是通过射频振荡时钟重新采样，因此整个系统使用的都是重新采样的时钟（CKR）。这就确保了大量的数字逻辑是在 TDC 检测相位误差的静默期之后时钟采样。

发射机架构是完全数字化的，并且通过调整数字频率指令来利用全数字锁相环的宽带调制能力。这种宽带调制方法完全是一种数字两点方案，一点是直接调制 DCO 频率导数，而另一点是补偿过多的相位误差。通过使数字逻辑提供的发射波形的失真尽可能最低，保持 DCO 增益特征稳定校准。幅度调制路径是通过工作在准 E 级模式的数字控制功率放大器（DPA）和大的 MOS 开关阵列实现的。射频幅度是通过打开的开关数量来控制。精确的幅度分辨率是通过高速晶体管开关振荡来实现。虽然数字逻辑运行速度很高，整个发射机架构的功耗比目前发射机架构的功耗要低。

对于 DRP 极性环路发射机和传统发射机也存在着折中。

1）把调制通路分成幅度调制和相位调制，从而避免使用混频器、滤波器、

甚至 D/A 转换器，但是不同通路可能出现延迟，为了达到高分辨率，高速调制器可能会增加功率耗散。

2）虽然 Σ-Δ 技术可以用于解决快速分数比特计算，相应地需要更精确的可变电容和线性，因此需要其他步骤，例如重新定时时钟、动态元件匹配（DEM）。

到目前为止，实际的 DRP 已经可以通过直接用数字电路取代模拟电路来实现。如图 2-14 所示，根据通信理论和基带通信设计开发更好的数字设计仍然是一个需要研究的问题，数字片上系统遵循摩尔定律，但是片上系统的模拟部分则不服从摩尔定律。

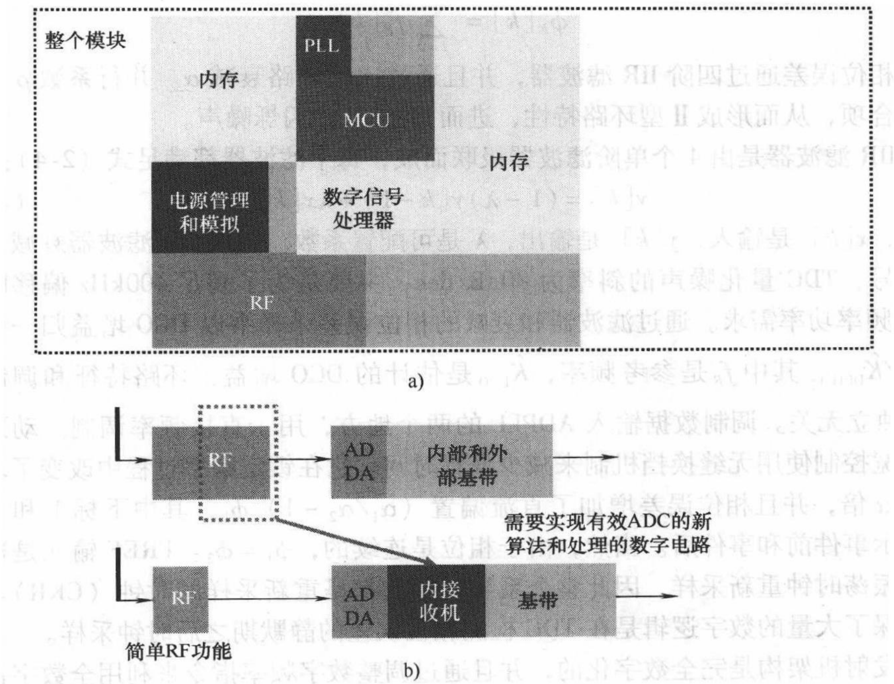


图 2-14 数字无线电处理

参考文献

[1] J. Mitola, III, 'Software Radio Architecture: A Mathematical Perspective', *IEEE Journal on Selected Areas in Communications*, **17**(4), 1999, 514–538.

[2] J.E. Gunn, K.S. Barron, W. Ruczczyk, 'A Low-power DSP Core-based Software Radio Architecture', *IEEE Journal on Selected Areas in Communications*, **17**(4), 1999, 574–590.

[3] U. Ramacher, 'Software-Defined Radio Prospects for Multistandard Mobile Phones', *IEEE Computer*, October 2007, 62–69.

[4] Yuan Lin, *et al.*, 'SODA: A High-Performance DSP Architecture for Software-Defined Radio', *IEEE Micro*, January-February 2007, 114–123.
[5] K.C. Chen, S.T. Wu, 'A Programmable Architecture for OFDM-CDMA', *IEEE Communications Magazine*, feature subject on Software and DSP in Radio, November 1999, 76–78.
[6] G. Gerrari, G. Colavolpe, R. Raheli, *Detection Algorithms for Wireless Communications*, John Wiley & Sons, Ltd, Chichester, 2004.
[7] H. Meyr, M. Moeneclaey, S. Fechtel, *Digital Communication Receivers*, John Wiley & Sons, Inc., New Jersey, 1998.
[8] C. Ebeling, *et al.*, 'Implementation an OFDM Receiver on RaPiD Reconfigurable Architecture', *IEEE Transactions on Computers*, **53**(11), 2004, 1436–1448.
[9] N. Yee, J.P. Linnartr, G. Fettweis, 'Multicarrier CDMA in Indoor Wireless Radio Networks,' *Proc. IEEE PIMRC '93, Yokohama, Japan*, September 1993, pp. 109–113.
[10] K. Fazel, L. Papke, 'On the Performance of Convolutionally-coded CDMA/OFDM for Mobile Communication System', *Proc IEEE PJMRC '93, Yokohama, Japan*, September 1993, pp. 468–472.
[11] V.M. DaSilva, E.S. Sousa, 'Performance of Orthogonal CDMA Codes for Quasi-Synchronous Communication Systems', *Proc. IEEE ICUPC*, Ottawa, Canada, 1993.
[12] L. Vandendorpe, 'Multitone Direct Sequence CDMA System in an Indoor Wireless Environment', *Proc. 1st JEEE Symp. Commun.*
[13] A. Loke, F. Ali, 'Direct conversion radio for digital mobile phones – Design issues, status and trends', *IEEE Trans. Microwave Theory Tech.*, **50**, 2002, 2422–2435.
[14] K. Muhammad, R.B. Staszewski, D. Leipold, 'Digital RF Processing: Towards Low-cost Reconfigurable Radios', *IEEE Communications Magazine*, August 2005, 105–113.
[15] K. Muhammad *et al.*, 'The First Fully Integrated Quad-band GSM/GPRS Receiver in a 90-nm Digital CMOS Process', *IEEE J. of Solid-State Circuits*, August 2006,
[16] R. B. Staszewski *et al.*, 'All-digital PLL and GSM/EDGE Transmitter in 90nm CMOS', *Proc. IEEE Solid-State Circuits Conf.*, Sec. 17.5, 600, February 2005, pp. 316–317.
[17] T. Sowlati, D. Rozenblit, R. Pullela *et al.*, 'Quad-band GSM/GPRS/EDGE polar loop transmitter', *IEEE J. Solid-State Circuits*, **39**(12), 2004, 2179–2189.
[18] G. Fettweis *et al.*, 'Dirty RF: A New Paradigm,' *International Journal of Wireless Information Networks*, **14**(2), 2007.
[19] J. Smith, *Modern Communication Circuits*, McGraw-Hill, New York, 1986.
[20] I. Martinez G., 'Automatic Gain Control (AGC) Circuits: Theory and Design,' University of Toronto 2001.
[21] F. Chen, B. Leung, 'A 0.25-mW Low-pass Passive Sigma-delta Modulator with Built-in Mixer for a 10-MHz IF Input,' *IEEE J. Solid-State Circuits*, **32**, June 1997, 774–782.
[22] V. Jimenez, J. Garcia, F. Serrano *et al.*, 'Design and Implementation of Synchronization and AGC for OFDM-based WLAN Receivers,' *IEEE Transactions on Consumer Electronics*, **50**(4), 2004, 1016–1025.
[23] I.-Gu Lee, S.-Kyu, 'Efficient Automatic Gain Control Algorithm and Architecture for Wireless LAN Receiver', *Journal of Systems Architecture*, **53**, 2007, 379–385.

第3章　无线网络

在介绍了无线通信物理层和软件无线电之后，本章我们将面向计算机网络高层协议的主要概念，也就是无线网络的基础。现代无线网络或移动网络可以分为两大类：基础设施和Ad hoc网络。

每个基础设施无线网络（见图3-1a）有一个高速骨干网（一般是有线网络）连接一些基站（或接入点）。移动台与骨干网络通过基站通信，并且进一步连接到目的移动台。分组数据的传递主要是依赖于基础设施，包括骨干网和基站。另一方面，一些移动台可能不需要基础设施而建立起Ad hoc网络，图3-1b中标明了两个节点（也就是移动台）之间的链路，这些链路也就构成了Ad hoc网络的网络拓扑。大多数现在的商业无线网络都是基于网络基础设施结构，而Ad hoc网络偶尔被允许使用。Ad hoc网络已经在军事系统中使用了很长一段时间，例如多跳分组无线网络。正如传感器等无线设备近年来具有很广阔的应用一样，Ad hoc网络将在无线网络和认知无线电网络中起到更重要的作用。

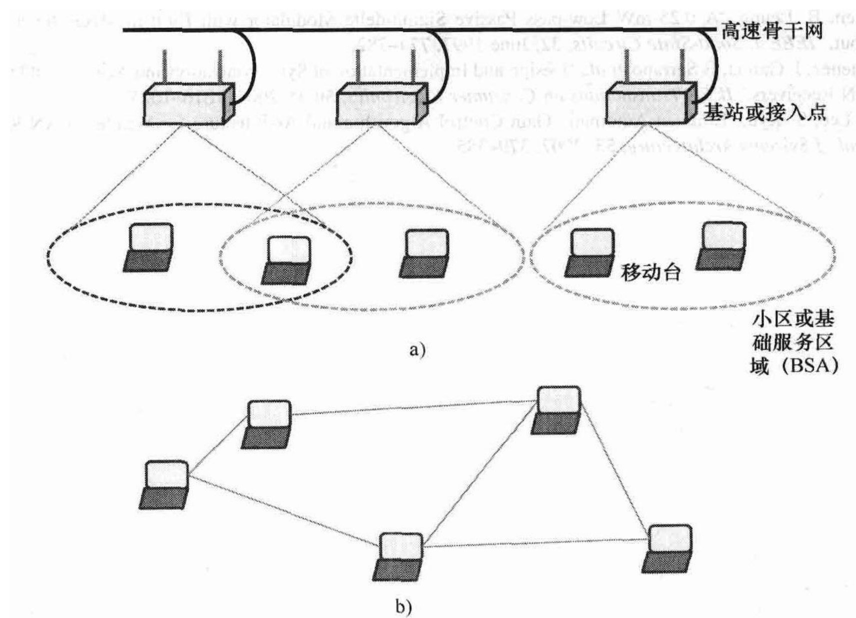


图3-1　a）基础设施和b）Ad hoc无线网络

3.1 多址接入通信和 ALOHA

回顾在第1章介绍过的网络分层结构，我们需要在数据链路层和物理层之间增加一个媒体接入控制（MAC）子层。增加这个子层的目的是为不同的节点分配多址接入媒体。在计算机/通信网络中的不同节点间协调物理传输的方法称为多址接入协议，这也是无线网络的关键功能。

多址接入协议工作在不同的网络拓扑和应用场景下。图3-2a给出了一个卫星通信网络的例子，地球站通过卫星相互通信，这是一个星形网络拓扑。图3-2b是一个总线网络拓扑，其中节点/站通过总线连接到网络（一般总线采用一根线缆或光纤作为物理传输媒介）。图3-2b中的另一个例子是一个多跳分组无线网络，一个Ad hoc无线网络中广泛采用的网络拓扑（有时使用网格网络扩展结构）和无线传感器网络。最早大家熟知的计算机网络是连接夏威夷各岛之间的卫星通信网络，下一节将介绍其中采用的多址接入协议ALOHA作为最早的多址接入协议。

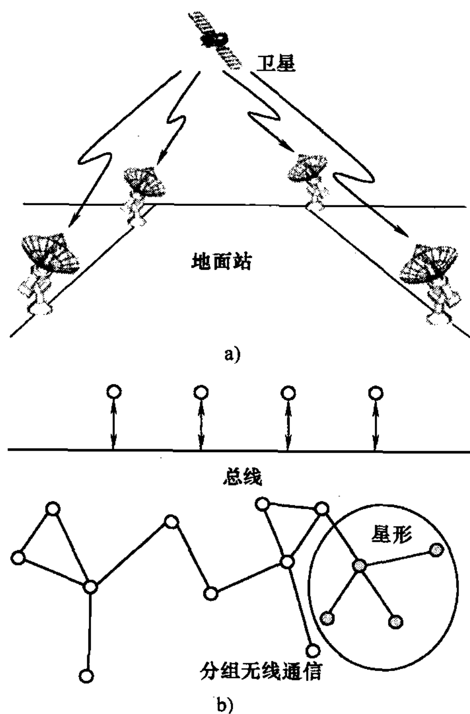


图3-2 多址接入通信的例子
a）卫星和地面站（星形拓扑，也是最简单的基础设施网络架构） b）总线、星形和多跳分组无线通信

3.1.1 ALOHA 系统和分时隙多址接入

最早的多址接入系统可能是用于卫星通信（实际上收集信息）系统的ALOHA系统，其拓扑如图3-2a所示。地面上的节点（地球站）试图接入卫星来中继传输数据包，因此需要一个多址协议来分布式协调各个节点的传输。当一些节点同时共享通信信道时，如果两个或更多节点同时传输时，接收会出现错误，也称为碰撞。此外，如果没有节点传输数据时，该信道是空闲的。多址接入就是为了解决如何通过分布式或集中式的方法来协调使用共享信道。在本节中，我们主要讨论分布式多址接入协议，分布式多址接入在无线数据网络中得到了广泛使

用。纯 ALOHA 协议十分简单：①当一个节点需要传输一个数据包，该节点传输数据。②该节点监听信道，如果发生了碰撞，该节点通过随机退避算法重新调度分组传输。否则，该节点成功传输了数据包。

为了研究多址接入协议，我们一般把时间轴分成时隙。理想的分时隙多址接入模型的假设可以归纳如下：

1）分时隙系统：所有要传输的数据包具有同样的长度，每个数据包需要一个时间单元（称为一个时隙）来传输。所有发射机都保持同步，因此每个分组数据包接收时都从整数时间开始，并且在下一个整数时间结束接收。

2）泊松到达：m 个传输节点传输的数据包到达服从到达速率 λ/m 的独立泊松过程。

3）碰撞或准确接收：分组数据接收时要么是由于碰撞丢失了信息，要么是准确地接收。

4）$\{0, 1, e\}$ 立即反馈：多址接入信道可以为分布式节点提供反馈，反馈的字母集为$\{0, 1, e\}$，其中 1 表示数据传输成功并准确接收，0 表示信道空闲没有数据传输，e 代表多址接入信道出现碰撞。

5）重传碰撞：每个出现碰撞的数据包都必须在后面的时隙内重传，并且可能还会继续重传，直到成功传输。具有重传数据包的节点称为出现积压。

6）①每个节点都没有缓存。②系统中节点数量无限。

众所周知，纯 ALOHA 网络的吞吐量（也就是在每个分组传输时间内平均成功传输的分组数）为 $1/(2e)$。

3.1.2 分时隙 ALOHA

这种方法的基本思路是每个未出现积压的节点在分组到达之后的第一个时隙内传输一个新到达的分组，因此可能会出现碰撞，但是如果碰撞较少时，该分组的延迟很小。对于 m 个节点，TDM 中到达的分组必须等待平均 $m/2$ 个时隙才能传输。分时隙 ALOHA 几乎立刻传输分组数据，虽然偶尔会出现碰撞，而 TDM 用更大的传输时延来避免碰撞。当分时隙 ALOHA 出现碰撞时，每个在碰撞出现时隙时传输碰撞分组的节点都将出现积压，这些节点等待随机个时隙以重传。由于假设节点数无限，一个时隙内新到达的分组数服从参数为 λ 的泊松随机变量。如果积压节点的重传足够随机化，那么可以近似把重传总的次数和给定时隙内新传输分组建模成参数 $G>\lambda$ 的泊松随机变量。一个时隙内成功传输的概率为 Ge^{-G}（见图 3-3）。

为了构建一个更精确的模型以理解动态系统，假设每个积压节点以同样的固定概率 q_r 在每个连续的时隙重传数据包直到传输成功。换句话说，从碰撞开始直到碰撞中的节点重传的时隙数是一个随机变量 $q_r(1-q_r)^{i-1}$，$i\geq 1$。我们根据

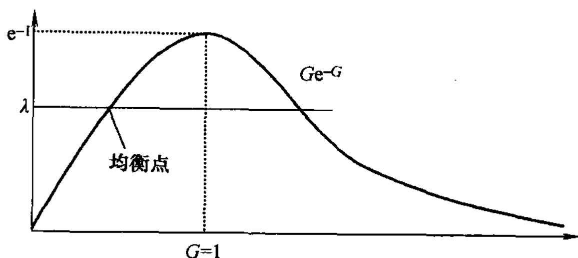


图3-3 传输参数

假设节点没有缓存，因此分时隙 ALOHA 可以描述成节点在给定时隙的开始时刻。每个节点在给定的时隙内相互独立地以概率 q_r传输数据包。如果在之前的时隙内有一个或多个数据分组到达，其他 $m-n$ 个节点将在给定的时隙内传输分组。由于数据的到达服从均值为 λ/m 的泊松过程，无分组到达的概率为 $e^{\lambda/m}$。因此，未积压节点在给定时隙传输分组的概率为 $q_a=1-e^{-\lambda/m}$。令 $Q_a(i,\ n)$ 是 i 个未积压节点在给定时隙传输数据包的概率，$Q_r(i,\ n)$ 是 i 个积压节点传输的概率：

$$Q_a(i,n)=\binom{m-n}{i}(1-q_a)^{m-n-i}q_a^i \tag{3-1}$$

$$Q_r(i,n)=\binom{n}{i}(1-q_r)^{n-i}q_r^i \tag{3-2}$$

从一个时隙到下一个时隙，马尔科夫链状态（积压分组的数量）随着未积压节点传输的新到达数据包的增加而增大，如果有数据包成功传输，那么状态将减1。一个数据包只有在下面条件满足的情况下才能成功传输：

1）一个新的到达数据包，没有积压的数据包。

2）没有新到达数据包，有一个积压的数据包。

图3-4 中从状态 n 到 $n+1$ 的状态转移概率是

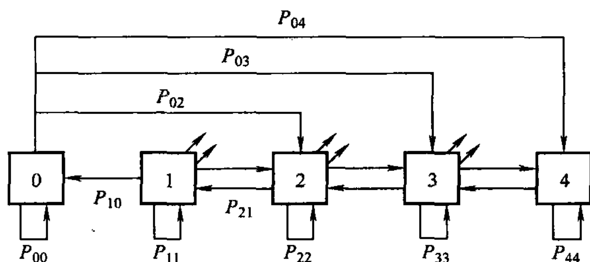


图3-4 分时隙 ALOHA 的马尔科夫链模型

$$P_{n,n+i}=\begin{cases}Q_a(i,n), & 2\leqslant i\leqslant (m-n)\\ Q_a(1,n)[1-Q_r(0,n)], & i=1\\ Q_a(1,n)Q_r(0,n)+Q_a(0,n)[1-Q_r(1,n)], i=0 & \\ Q_a(0,n)Q_r(1,n), i=-1 & \end{cases} \tag{3-3}$$

我们可以选择足够大的重传概率 q_r，以避免过大的延迟，$q_r n\gg 1$ 是可能存在的，但是这样会导致几乎所有连续的时隙都会出现碰撞，最终系统将出现长期严重的积压。

令 D_n是在一个时隙内从状态 n 开始预计变化的次数，P_{SUCC}表示成功传输的概率，那么预计成功传输的分组数为

$$D_n=(m-n)q_a-P_{\text{SUCC}} \tag{3-4}$$

其中

$$P_{\text{SUCC}}=Q_a(1,n)Q_r(0,n)+Q_a(0,n)Q_r(1,n) \tag{3-5}$$

我们定义尝试速率 $G(n)$ 表示当系统处于状态 n 时一个时隙内预计尝试传输的次数：

$$G(n)=(m-n)q_a+nq_r \tag{3-6}$$

如果 q_a、q_r足够小，

$$P_{\text{SUCC}}\approx G(n)\mathrm{e}^{-G(n)} \tag{3-7}$$

出现空闲时隙的概率大约是 $\mathrm{e}^{-G(n)}$。

如图 3-5 所示，在曲线和直线之间存在着差异，由于偏移表示从一个时隙到下一个时隙状态预计的变化，虽然系统会出现波动，但是系统仍然会往偏移的趋势发展，最终聚集在两个稳定点，并且这两个稳定点之间相差很少。

由上可得以下的结论：

1）在 m 较大时，P_{SUCC}离开率最大能达到 $1/\mathrm{e}$。

2）无论何时系统跳转到不想达到的稳定点，从长期来看，离开率几乎为 0。

考虑到 q_r的变化，随着 q_r的增大，重传碰撞数据包的延迟减少，但是 n 和尝试速率 $G(n)$ 之间的线性关系也出现了变化。如果 n 的水平比例保持恒定，尝试速率的变化将缩小 $G(n)$ 的水平比例，因此偏离 Ge^{-G}曲线。也就意味着超过不稳定均衡点所需的积压分组数量减少。如果 q_r减少足够大（仍然大于 q_a），曲线 Ge^{-G}将进一步扩展，有且仅有一个稳定状态。在这个稳定点，当 $q_r=q_a$时，积压的分组数是 m 的相当一部分（见图 3-5）。这就表示相当多的到达数据包会被丢弃，因此延迟会非常大。

3.1.3 稳定分时隙 ALOHA

ALOHA（与之后其他章节的 ALOHA 系列）是不稳定的随机接入协议，也

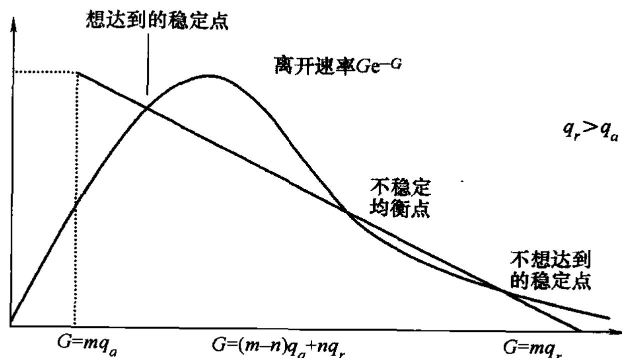


图3-5 ALOHA吞吐量曲线的稳定状态

就意味着当到达的业务量超过稳定范围时，网络吞吐量为0（也就是延迟无限大），最后导致全部业务数据包都积压起来。

为了保持高的多址信道效率，需要动态改变q_r来保持$G(n)$为1，从而最大化$P_{\text{SUCC}} \approx G(n)e^{-G(n)}$。问题在于$n$对于节点来说是未知数，只能通过反馈来估计$n$。所有估计$n$或者估计$q_r$的方法本质上都是在出现空闲时隙时增大$q_r$，而在出现碰撞时减少$q_r$。

假设节点数无限，随着系统继续运行，该系统并没有稳定的状态分布，预期的网络延迟也会无限增大。假设节点没有缓存，系统丢弃了大量的到达数据包，因此网络延迟很大但是有限，而由于假设节点无限，将不会丢弃到达的数据，但是网络延迟会无限大。

多址接入系统的最大稳定吞吐量定义成系统稳定时最小的到达速率的上限。例如对分时隙ALOHA，其最大稳定吞吐量为0。

当假设积压的估计准确时，$G(n)$最优值保持为1，空闲出现的概率为$1/e \approx 0.368$，成功传输的概率为$1/e$，碰撞出现的概率为$1-2/e \approx 0.264$。因此，改变q_r的规则应该允许碰撞比空闲出现得更少。这种系统的最大稳定吞吐量最多能达到$1/e$。

伪贝叶斯算法可能是最著名的稳定ALOHA算法和协议。该算法与分时隙ALOHA的不同在于新到达数据包一旦到达就当做积压数据包处理，在分时隙ALOHA中心到达的数据包会在下一个时隙内传输，而在伪贝叶斯算法中则会与之前碰撞中的分组一样，以概率q_r传输。如果在一个时隙的开始时刻有n个积压的数据包（包括新到达的数据包），尝试率为$G(n)=nq_r$，成功传输的概率为$nq_r(1-q_r)^{n-1}$。对于不稳定ALOHA，q_r必须相对较小，这样新到达的数据就不会出现不必要的延迟。而对于稳定ALOHA系统，当可以忽略估计的积压数据包时，q_r可以取较大的值，甚至取1，这样当估计系统已经出现拥塞时，新到达的

数据包将会被丢弃。

该算法是通过在每个时隙开始时估计积压的数据包数量 n，估计值假设为$\hat{n}$。每个积压的数据包以概率 $q_r=\min\{1, 1/\hat{n}\}$ 传输（独立的），其中，q_r最大为 1，从而达到尝试率 $G(n)=nq_r$。

$$\hat{n}_{k+1}=\begin{cases}\max\{\lambda,\hat{n}_k+\lambda-1\}，对于空闲时隙或成功传输\\ \hat{n}_k+\lambda+(e-2)^{-1}，对于碰撞\end{cases} \tag{3-8}$$

3.1.4 分析近似时延

假设已知 λ，当 $n\geqslant 2$ 时，成功传输的概率 $P_{\mathrm{SUCC}}=1/e$ ，当 n 为 1 时，$P_{\mathrm{SUCC}}=1$。对于 λ 较小时，这是一个合理的模型，因为很少有碰撞发生，q_r一般为 1。对于 $\lambda>1/e$ 时也很合理，因为一般 n 很大，$\hat{n}\approx n$。

令 W_i表示从第 i 个数据包到达到第 i 次成功传输时的时延。W_i对 i 取平均等于期望的队列延迟 W。令 n_i表示在 i 到达时刻之前积压的数据包数量。N_i不包括目前正在被成功传输的数据包，但是包括目前没有成功传输的数据包。

$$W_i=R_i+\sum_{j=1}^{n_i}t_{j_i}+y_i \tag{3-9}$$

式中，R_i是到下一个时隙开始时刻剩余的时间，$t_1(n_i>0)$ 表示到下一次成功传输完成的时间间隔，$t_j(1<j\leqslant n_i)$ 表示第 $j-1$ 次成功传输到第 j 次成功传输之间的间隔。在 n_i次成功传输之后，y_i是到下一次成功传输开始时刻的剩余间隔。

每个时隙成功传输的概率为 1/e，t_j的期望值为 e，根据 Little 定律，

$$W=\frac{1}{2}+\lambda eW+E\{y\} \tag{3-10}$$

考虑系统在第一个时隙的边界时，第 $i-1$ 次数据包离开，第 i 个数据包到达。如果积压为 1，那么 $y_i=0$。如果 $n>1$，$E\{y_i\}=e-1$。令 P_n表示在时隙边界时积压为 n 的稳定状态概率。如果状态为 1，那么数据包总是成功传输，P_1是状态为 1，并且数据包成功传输的部分时隙。由于 λ 是总的成功传输时隙的比例，P_1/λ 是从状态 1 传输的数据包比例，$1-P_1/\lambda$ 是从更大序号状态传输数据包的比例：

$$E\{y\}=\frac{(e-1)(\lambda-P_1)}{\lambda} \tag{3-11}$$

$$\lambda=P_1+(1-P_0-P_1)/e \tag{3-12}$$

在时隙边界，当且仅当在前一个时隙没有数据包到达，并且之前的状态为 0 或者 1 时才能进入状态 0，因此：

$$P_0=(P_0+P_1)e^{-\lambda} \tag{3-13}$$

$$P_1 = \frac{(1-\lambda e)(e^{\lambda}-1)}{1-(e-1)(e^{\lambda}-1)} \tag{3-14}$$

$$W = \frac{e-\frac{1}{2}}{1-\lambda e} - \frac{(e^{\lambda}-1)(e-1)}{\lambda[1-(e-1)(e^{\lambda}-1)]} \tag{3-15}$$

因此，我们可以用分析的方法得到等待时间（也就是延迟）。

3.1.5 未分时隙 ALOHA

在纯 ALOHA 方案中没有分时隙的概念，因此可以认为是多址接入协议最原始的版本。一个需要传输数据包的节点传输数据包并监听信道，如果一个数据包出现碰撞，那么该节点在随机延迟之后重新传输该数据包。

从图 3-6 可以看出，碰撞持续时间最多是分时隙 ALOHA 系统中碰撞持续时间的两倍。因此，如图 3-7 所示，

$$吞吐量 = G(n)e^{-2G(n)}$$



图 3-6 碰撞中的数据包

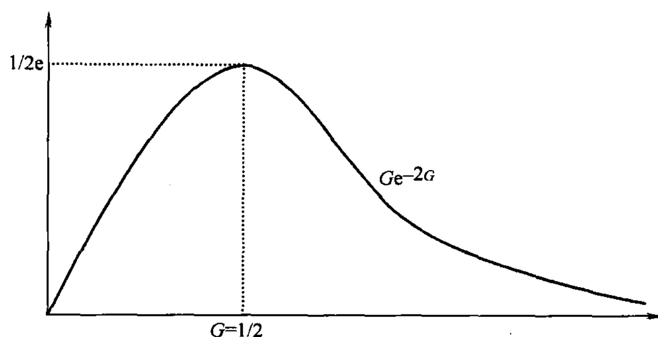


图 3-7 未分时隙 ALOHA 吞吐量

读者可以参考文献［1，2］对 ALOHA 系列协议延迟的详细分析。

3.2 分裂算法

ALOHA 协议系列中很重要的一个种类——随机退避算法用于解决碰撞问题。一般都通过成熟的碰撞解决技术来维护网络稳定，而避免使用复杂的估计过程，同时还能增加网络吞吐量。直观上说，使用较小的尝试率，碰撞最可能发生在两个数据包之间。如果禁止在碰撞解决之前传输新到达的数据包，那么每个碰撞的

数据包可以在下一个时隙以 1/2 的概率独立重传，重复这个成功传输，然后再传输其他数据包。

发送这样两个数据包预期的时隙数是 3，在解决碰撞期间的吞吐量为 2/3。碰撞中的节点可以用不同的方法来选择是否在下面的时隙传输：

1）抛硬币随机决定。

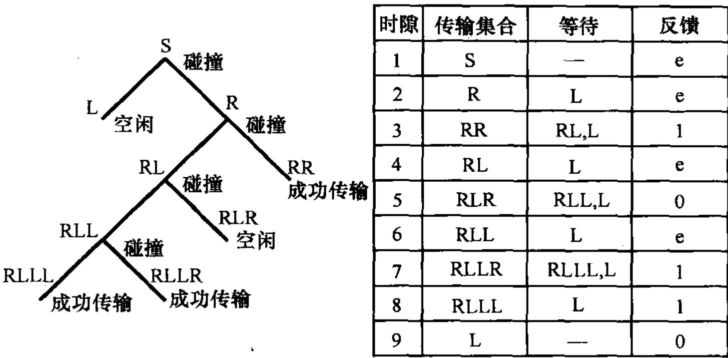
2）使用碰撞数据包的到达时间。

3）假设节点数有限，每个节点有惟一的用比特串表示的标识，因此节点可以使用该连续标识比特来连续作出选择，其优点在于只需要有限的时隙来解决碰撞。

我们称这类算法为分裂算法或者解决碰撞协议（CRP）。

3.2.1 树算法

当在第 k 个时隙内发生碰撞时，所有碰撞中的节点进入等待模式，所有涉及碰撞的节点分为两个子类（例如，通过抛硬币决定）。第一个子集在第 $k+1$ 个时隙传输，如果该时隙空闲或者成功传输，第二个子集在第 $k+2$ 个时隙传输。或者，如果在第 $k+1$ 个时隙发生另一起碰撞，第二个子集的前一个子集继续分裂，后面的子集继续等待碰撞解决（见图 3-8）。



时隙	传输集合	等待	反馈
1	S	—	e
2	R	L	e
3	RR	RL,L	1
4	RL	L	e
5	RLR	RLL,L	0
6	RLL	L	e
7	RLLR	RLLL,L	1
8	RLLL	L	1
9	L	—	0

图 3-8 树算法

一个问题在于当正在解决碰撞时，该如何处理新到达的数据包。碰撞解决周期（CRP）在成功传输或者出现空闲时隙时结束，并且此时堆栈里没有剩余的数据包。此时，新的 CRP 开始使用在之前 CRP 时间内到达的数据包。然而，仍然有新的等待到达数据包发生碰撞，并且持续碰撞直到该子集在该新的 CRP 内变得足够小。解决办法是在每个 CRP 时间的结尾处，有新到达数据的节点集立即分裂成 j 个子集，其中选择 j 为每个子集预期的数据包数，一般比 1 稍大。这些新子集进入堆栈，开始新的 CRP。最大吞吐量为 0.43。

如果碰撞之后是空闲时隙，就意味着所有发生碰撞的数据包都分配到了第二级子集。最大吞吐量是0.46。

实际中，这种改进方法有一些小问题，如果接收机把空闲时隙错误地当成发生了碰撞，该算法会继续无限分裂下去，从此不会再成功传输。因此，在分裂之后的第 h 个空闲时隙，该算法应该简单地作出修正，不再继续分裂，而是传输堆栈上的下一个子集。如果反馈信息很可靠，h 可以适量取较大的值。令 x 是第一次碰撞中数据包的数量，X_L和 X_R是在子集中的数据包数。$X = X_L + X_R$。假设预先知道 X 是泊松分布，那么 X_L和 X_R是独立的均值为 X 均值一半的泊松随机变量。假定 $X_L + X_R \geqslant 2$，并且 $X_R \geqslant 2$，那么将不把一个时隙分给第二级子集中预期数据包最少的子集，更好的办法是把第二级子集看成是从未发生碰撞的等待新到达数据包的另一部分节点。

3.2.2 FCFS 分裂算法

通过根据到达时间分裂，每个子集由在给定时段到达的所有数据包组成。当发生碰撞时，该时间段将继续分裂成两个小的时间段。在每个整数时间 k，该算法指定了在时隙 k 将传输在较早时间段到达的所有数据包集。该时间段称为时隙 k 的分配区间，$T(k)$到 $T(k)+\alpha(k)$。

每个节点根据精确的先到先服务（FCFS）算法如图3-9所示：

如果反馈=0，那么

$$T(k) = T(k-1)$$
$$\alpha(k) = \alpha(k-1)/2$$
$$\sigma(k) = \mathrm{L}$$

如果反馈=1，并且 $\sigma(k-1) = \mathrm{L}$，那么

$$T(k) = T(k-1) + \alpha(k-1)$$
$$\alpha(k) = \alpha(k-1)$$
$$\sigma(k) = \mathrm{R}$$

如果反馈=0，并且 $\sigma(k-1) = \mathrm{L}$，那么

$$T(k) = T(k-1) + \alpha(k-1)$$
$$\alpha(k) = \alpha(k-1)/2$$
$$\sigma(k) = \mathrm{R}$$

如果反馈=0或1，并且 $\sigma(k-1) = \mathrm{R}$，那么

$$T(k) = T(k-1) + \alpha(k-1)$$
$$\alpha(k) = \min[\alpha_0, k - T(k)]$$
$$\sigma(k) = \mathrm{L}$$

其中，可以选择 α_0使给定到达速率时延迟最小或者稳定吞吐量最大。

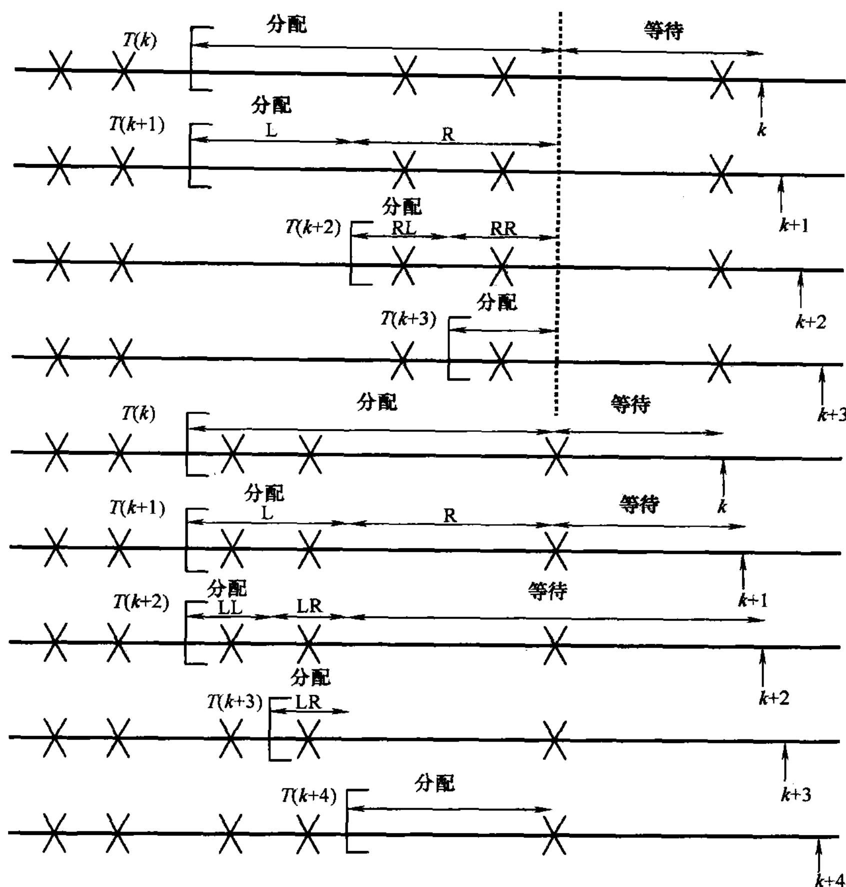


图 3-9 FCFS 算法

该算法给出了分配区间（也就是 $T(k)\alpha(k)$），根据反馈、分配区间和时隙 $k-1$ 状态 $\sigma=$（L 或 R）来确定时隙 k 的状态。

3.2.3 分析 FCFS 分裂算法

节点（R，0）对应的碰撞解决周期 CRP 的初始时隙分裂成两部分，从而使 CRP 的开始和结束时刻更加清楚。

假设初始分配间隔大小为 α_0，每次分裂时分配间隔分裂成两部分，图 3-10 中的 (L,i) 和 (R,i) 对应大小为 $2^{-i}\lambda\alpha_0$ 的分配间隔。定义 G_i 为已经分裂 i 次的时间间隔内预期到达的数据包数：

$$G_i=2^{-i}\lambda\alpha_0 \tag{3-16}$$

图 3-10 可以看成是一个马尔科夫链，为了找到转移概率，我们只考虑其中一个碰撞解决周期。节点 $(R,0)$ 的上半部分是开始状态，而下半部分是最终状

态。$P_{(R,0)}$是在第一个时隙内空闲或成功传输的概率。由于初始分配间隔内数据包数是均值为G_0的泊松分布：

$$P_{(R,0)}=(1+G_0)e^{-G_0} \tag{3-17}$$

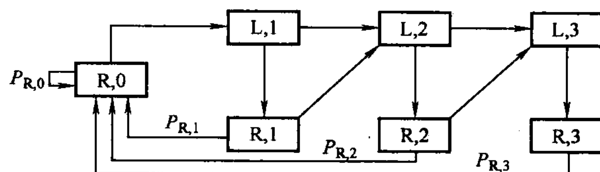


图3-10 FCFS分裂算法

(R,0)在出现碰撞之后以概率$1-P_{(R,0)}$进入(L,1)，X_L和X_R分别是在新的分配间隔L和R中数据包的数量，它们是独立服从均值G_1的泊松随机变量。

$$P_{(L,1)}=\frac{P\{X_L\}P\{X_R\geqslant 1\}}{P(X_L+X_R\geqslant 2)}=\frac{G_1e^{-G_1}(1-e^{-G_1})}{1-(1+G_0)e^{-G_0}} \tag{3-18}$$

当且仅当在上面的转换之后才会进入（R，1）。因此，成功进入的概率$P_{R,1}$是

$$P_{(R,1)}=\frac{P\{X_R=1\}}{P\{X_R\geqslant 1\}}=\frac{G_1e^{-G_1}}{1-e^{-G_1}} \tag{3-19}$$

可以证明

$$P_{(L,i)}=\frac{G_ie^{-G_i}(1-e^{-G_i})}{1-(1+G_{i-1})e^{-G_{i-1}}} \tag{3-20}$$

$$P_{(R,i)}=\frac{G_ie^{-G_i}}{1-e^{-G_i}} \tag{3-21}$$

在返回（R,0）之前进入（L,i)(R,i）状态的概率可以根据（R,0）迭代计算为

$$P(L,1)=1-P_{R,0} \tag{3-22}$$

$$p(R,i)=P_{(L,i)}p(L,i),I\geqslant 1 \tag{3-23}$$

$$P(L,I+1)=(1-P_{L,i})p(L,i)+(1-P_{(R,i)})p(R,i),I\geqslant 1 \tag{3-24}$$

令k表示CRP中的时隙数，因此，k是马尔科夫链访问的状态数量，包括在返回（R，0）状态之前的（R，0）

$$E\{k\} = 1+\sum_{i=1}^{\infty}[P(L,i)+P(R,i)] \tag{3-25}$$

对于设定的α_0，$T(k)$最多改变α_0次。但是如果左边间隔返回到等待间隔，$T(k)$的改变少于α_0。令f表示在一个CRP内按这种方法返回的α_0比例，因此$T(k)$改变了$\alpha_0(1-f)$。状态（L,i）碰撞的概率是（L,i）中左边间隔中包括至少两个数据包，并且右边和左边间隔一共包含至少两个数据包的概率。

$$P\{e\mid (L,i)\}=\frac{1-(1+G_i)e^{-G_i}}{1-(1+G_{i-1})e^{-G_{i-1}}} \tag{3-26}$$

初始间隔返回碰撞的比例是 2^{-i}

$$E\{f\} = \sum_{i=1}^{\infty} p(\mathrm{L},i)P\{\mathrm{e} \mid (\mathrm{L},i)\}2^{-i} \tag{3-27}$$

$E\{f\}$ 和 $E\{k\}$ 只是 $G_i = \lambda\alpha_0 2^{-i}$ 的函数，因此也只是乘积 $\lambda\alpha_0$ 的函数。当 i 较大时，$P_{\mathrm{L},i} \to 1/2$，$p(\mathrm{L},\ i)$ 和 $p(\mathrm{R},\ i)$ 趋向于 0。

定义偏移 D 为在一个 CRP 内时间积压 $k - T(k)$ 中预期的改变。也就是在一个 CRP 内少于 $T(k)$ 中预期改变的时隙数。

$$D = E\{k\} - \alpha_0(1 - E\{f\}) \tag{3-28}$$

当 $\mathrm{E}\{k\} < \alpha_0(1 - f)$ 时，该偏移是非负的，上面的条件可以等效为

$$\lambda < \frac{\lambda\alpha_0(1 - E\{f\})}{E\{k\}} \tag{3-29}$$

$$E\{k\} = 0.4971, \text{ 当 } \lambda\alpha_0 = 1.206 \text{ 时} \tag{3-30}$$

也就意味着 FCFS CRP 算法能达到的最大吞吐量是 0.4871，S. Verdu 稍后更精确地计算出最大吞吐量为 0.48760。

3.3 载波监听

事实上，如果节点能够从多址接入信道收集到一些信息，则可以设计出具有更优性能的多址接入协议。最直接的方法可能就是载波监听，节点首先监听要传输的信道，然后再进行传输。因此，载波监听也称为传输前监听（LBT）。

为了建模载波监听，令 β 表示所有源节点在传输结束之后检测空闲信道需要的归一化传输和检测时延：

$$\beta = \frac{\tau c}{L} \tag{3-31}$$

式中，τ 是时间（s），c 是原始信道比特速率，L 是数据包的比特数。

3.3.1 CSMA 分时隙 ALOHA

最简单的改进 ALOHA 方案的办法是引入载波监听多址接入（CSMA）概念。CSMA 分时隙 ALOHA 与普通的分时隙 ALOHA 的主要区别在于：

1）CSMA 的空闲时隙长度为 β；

2）如果在传输时有新的数据包到达节点，该数据包被认为是积压，并以概率 q_r 在每个空闲时隙传输；在空闲时隙内到达的数据包通常会在下一个时隙传输。

一般有三类载波监听方案：

1）非坚持：正如之前介绍的；

2）坚持（或者1-坚持）：所有在繁忙时隙内到达的数据包推迟到该时隙结束时传输；

3）p-坚持：碰撞的数据包和新到达的数据包等到繁忙时期结束后使用不同的概率传输。

在非坚持CSMA方案中，用户产生一个数据包，发现信道繁忙而不能传输数据包，就和数据包发生碰撞一样，也就是随机调度在未来某个时刻重传数据包。注意非坚持CSMA等效于特定p值的p-坚持CSMA。

3.3.1.1 **吞吐量分析**

有如下假设：

1）初始用户产生的数据包服从参数为λ的泊松过程。

2）所有的数据包具有同样的长度，传输时间为T。在观察信道时，数据包到达（新到达和重传的）服从参数为g个数据包/s的泊松过程。

3）τ为最大传输时延。

$a=\tau/T$为归一化传输时间。

从图3-11可以看出沿着时间轴的连续周期，每个周期由一个传输期和一个空闲期组成。$\tilde{B}$是繁忙期的持续时间，均值为B；$\tilde{U}$是传输期的持续时间，在传输期内成功地传输了数据包，均值为U；$\tilde{I}$是均值为I的空闲期持续时间。

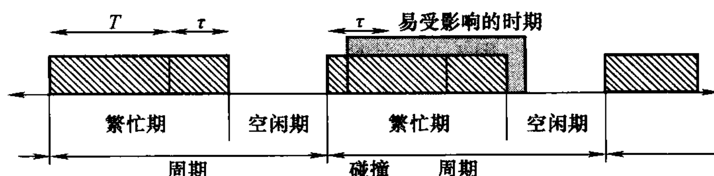


图3-11 非坚持CSMA

整个周期的长度是$\tilde{B}+\tilde{I}$，吞吐量是$S=U/(B+I)$。

空闲期的持续时间等于传输结束到新数据包到达之间的时间。由于数据包调度是无记忆的，

$$\begin{aligned}F_I(x)&=P_r[\tilde{I}\leqslant x]=1-P_r[\tilde{I}>x]\\&=1-P_r[x]\\&=1-e^{-gx}\end{aligned}\tag{3-32}$$

式中，在x期间没有可调度的数据包。

$\tilde{I}$服从均值$I=1/g$的指数分布。

为了得到 U，

$$\tilde{U}=\begin{cases}T,\text{成功传输期}\\0,\text{失败传输期}\end{cases}\tag{3-33}$$

如果 P_{SUCC}表示传输数据包成功的概率，那么

$$U=\mathrm{E}[\tilde{U}]=TP_{\text{SUCC}}\tag{3-34}$$

$$\begin{aligned}P_{\text{SUCC}}&=P_r[\text{在}[t,t+\tau]\text{期间没有数据到达}]\\&=\mathrm{e}^{-gt}\end{aligned}\tag{3-35}$$

那么

$$U=T\mathrm{e}^{-gt}\tag{3-36}$$

为了计算 B，令 $\tilde{Y}$ 是一个随机变量，因此 $t+\tilde{Y}$ 表示最后干扰的数据包在时间 t 开始的传输期内调度的时间（见图3-12）。

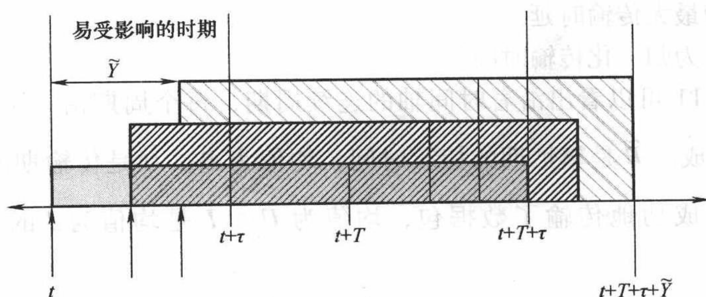


图3-12 计算 B 的示意图

很明显，对于成功传输期，$\tilde{Y}<\tau$，$\tilde{Y}=0$。

$$\tilde{B}=T+\tau+\tilde{Y}\tag{3-37}$$

周期 $\tilde{Y}$ 的特点是在［$t+\tilde{Y},t+\tau$］内没有其他数据包调度传输。否则，在 $t+\tilde{Y}$ 传输的数据包将不是在［$t+\tilde{Y},t+\tau$］内传输的最后一个数据包。

$$\begin{aligned}F_Y(y)&=P_r[\tilde{Y}\leqslant y]\\&=P_r[\text{在}\ \tau-y\ \text{之间没有数据包到达}]\\&=\begin{cases}\mathrm{e}^{-g(\tau-y)},0\leqslant y\leqslant\tau\\0,y<0\\1,y\geqslant\tau\end{cases}\end{aligned}\tag{3-38}$$

$$f_Y(y)=\mathrm{e}^{-gt}\delta(y)+g\mathrm{e}^{-g(\tau-y)}\tag{3-39}$$

$$\mathrm{E}[\tilde{Y}]=\tau-\frac{1-\mathrm{e}^{-gt}}{g}\tag{3-40}$$

因此

$$B = \mathrm{E}[T + \tau + \tilde{Y}] = T + 2\tau + \frac{1 - \mathrm{e}^{-gt}}{g} \tag{3-41}$$

$$S = \frac{gT\mathrm{e}^{-g\tau}}{g(T + 2\tau) + \mathrm{e}^{-gt}} \tag{3-42}$$

我们对分组传输时间归一化（见图 3-13），令 G 表示平均分组调度率，用每个分组传输时间内的分组数来衡量，也就是

$$G = gT \tag{3-43}$$

$$S = \frac{G\mathrm{e}^{-aG}}{G(1 + 2a) + \mathrm{e}^{-aG}} \tag{3-44}$$

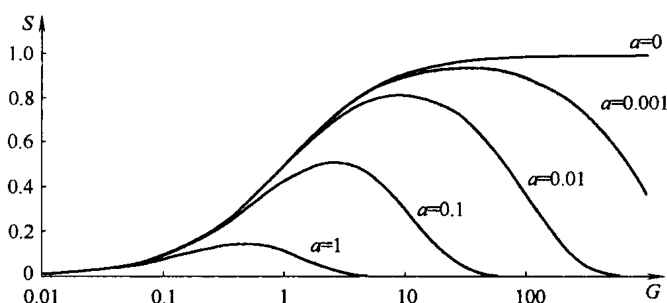


图 3-13　归一化吞吐量与分组传输时间

注意到当 $a \to 0$ 时，

$$S \to \frac{G}{(1 + G)}$$

随着负荷增大，吞吐量并不会减少到 0。最后，CSMA 可以看成是 ALOHA 系列协议中的一员，因此也不稳定。

3.3.1.2　1-坚持 CSMA

节点/用户监测信道，如果信道繁忙，该节点坚持等待，一旦信道空闲就开始传输。因此，如果有用户需要传输数据时，信道一直被占用（见图 3-14）。

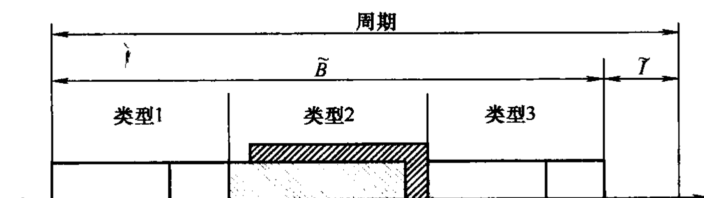


图 3-14　1-坚持 CSMA 原理

传输期要么是从一个数据包的传输开始（类型 1），要么从传输至少两个数据包开始（类型 2）。在空闲期后的传输期一般都是类型 1 传输期。空闲期称为类型 0。定义系统在传输期开始时的状态为传输期类型。这些状态对应于在传输期开始时的三状态马尔科夫链（见图 3-15）。

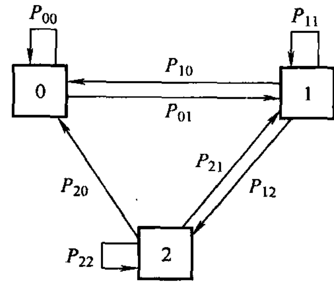


图 3-15　1-坚持 CSMA 的三状态马尔科夫链

只有类型 1 传输期能成功传输。然而，类型 1 传输期成功传输还需要在前 τ 秒内（易受影响期）没有数据包到达，每个数据包到达的概率为 e^{-gt}。

令 π_i表示在状态 i 的稳定概率。令 $\tilde{T}_i$（$i=0，1，2$）表示类型 i 传输期长度的随机变量，并且 $E[\tilde{T}_i]=T_i$。

$$S = T\frac{\pi e^{-gt}}{\sum_{i=0}^{2}\pi_i T_i} \tag{3-45}$$

1-坚持 CSMA 中的空闲期与非坚持 CSMA 相同，也就是均值为 $1/g$ 的指数分布。$\tilde{T}_1$和 $\tilde{T}_2$有同样的分布，因为一个数据包或两个以上数据包的传输期长度是由上一个数据包（如果有）在易受影响期内的到达时间决定的，而与传输期的类型无关。假设传输期从时间 t 开始，在时间 t 开始的传输期的易受影响期内到达的上一个数据包的到达时间（在 t 时刻之后）是一个随机变量，用 $\tilde{Y}$ 表示，如果在 $[t,\ t+\tau]$ 之内没有数据包到达：

$$\tilde{T}_1 = \tilde{T}_2 = T+\tau+\tilde{Y} \tag{3-46}$$

$$f_Y(y) = e^{-gt}\delta(y) + ge^{-g(\tau-y)} \tag{3-47}$$

$$E[\tilde{Y}] = \tau - \frac{1-e^{-gt}}{g} \tag{3-48}$$

$$T_1 = T_2 = T+\tau+E[\tilde{Y}] = T+\tau-\frac{1-e^{-gy}}{g} \tag{3-49}$$

从状态转移图可以得到：

$$\pi_0 = \pi_1 p_{10} + \pi_2 p_{20} \tag{3-50}$$

$$\pi_1 p_{12} = \pi_2(p_{21}+p_{20}) \tag{3-51}$$

$$\pi_0+\pi_1+\pi_2 = 1 \tag{3-52}$$

当类型 1 或类型 2 传输期开始后，下一个传输期的类型是由在传输期开始之后调度传输的数据包决定的。如果在传输期内没有数据包到达，那么下一个传输

期将是类型0。如果只有一个数据包到达，至少在传输期之后 τ 秒，下一个传输期将是类型1。如果在传输期内至少到达两个数据包，至少在传输期之后 τ 秒，下一个传输期将是类型2：

$$p_{1j} = p_{2j} \tag{3-53}$$

可以得到

$$\pi_0 = \frac{p_{10}}{1 + p_{10}} \quad \pi_1 = \frac{p_{10} + p_{11}}{1 + p_{10}} \quad \pi_2 = \frac{1 - p_{10} - p_{11}}{1 + p_{10}} \tag{3-54}$$

假设类型1传输期在时间 t 开始。当 $\tilde{Y} = y$ 时，只有当在 $[t+\tau, t+y+T+\tau]$ 内没有数据包调度时，下一个传输期将是类型0。没有数据包调度的概率为 $e^{-g(T+y)}$。

$$\begin{aligned} p_{10} &= \int_0^{\tau} e^{-g(T+y)} f_Y(y) \\ &= \int_0^{\tau} e^{-g(T+y)} [e^{-g\tau}\delta(y) + g e^{-g(\tau-y)}] dy \\ &= g e^{-g(T+\tau)} \left[T + g\tau\left(T + \frac{\tau}{2}\right)\right] \end{aligned} \tag{3-55}$$

可以得到

$$S = \frac{G e^{-G(1+2a)} \left[1 + G + aG\left(1 + G + \frac{aG}{2}\right)\right]}{G(1+2a) - (1 - e^{-aG}) + (1 + aG) e^{-G(1+a)}} \tag{3-56}$$

1-坚持 CSMA 的性能如图3-16所示。

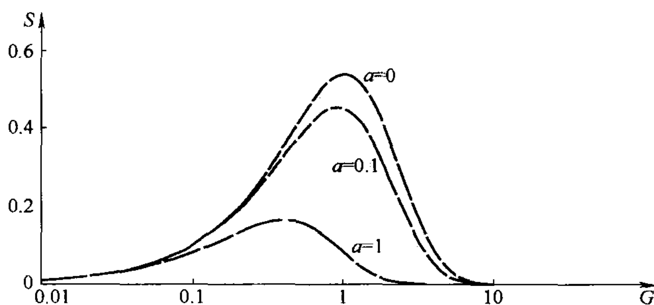


图3-16 1-坚持 CSMA 的性能

3.3.2 分时隙 CSMA

假设时隙长度为 τ，也就是任何在时隙起点开始的传输都是在时隙结束时到达（可以监测）。这些时隙可以看成是微时隙。用户被限制在微时隙的边界处开始传输。假设 τ 包括传输时延和载波监听的时间。还假设 T 是 τ 的整数倍。（$1/a$

是整数。)

当一个数据包在给定时间内调度传输，用户等到下一个微时隙的开始时刻，此时用户监测信道，如果信道空闲就传输该数据包。(该数据包在所有其他用户接收之前占据信道 $1/a+1$ 个微时隙) 如果信道监测繁忙，将执行相应的 CSMA 协议。

3.3.2.1 分时隙非坚持 CSMA 的吞吐量

如果在前 $k-1$ 个微时隙没有数据包到达，在第 k 个微时隙有至少一个数据包到达，那么空闲期包括 k 个微时隙：

$$p\{\tilde{I}=k\tau\}=(e^{-g\tau})^{k-1}(1-e^{-g\tau})\quad k=1,2,\cdots \tag{3-57}$$

$$I=\frac{\tau}{1-e^{-g\tau}} \tag{3-58}$$

成功和失败传输期都持续 $T+\tau$ 秒。如果两个或多个数据包在同一个微时隙内到达，并且都在下一个微时隙内调度传输，将会发生碰撞。如果在前 $k-1$ 个传输期的上一个微时隙内都至少有一个数据包到达，并且在第 k 个传输期的上一个微时隙内没有数据包到达，那么繁忙期将包括 k 个传输期：

$$P_r\{\tilde{B}=k(T+\tau)\}=(1-e^{-gt})^{k-1}e^{-gt}\quad k=1,2,\cdots \tag{3-59}$$

$$B=\frac{T+\tau}{e^{-gt}} \tag{3-60}$$

与前面的推导类似，

$$U=E\{\tilde{U}\}=T\frac{B}{T+\tau}P_{SUCC} \tag{3-61}$$

$$\begin{aligned}P_{SUCC}&=P_r\{成功传输期\}\\&=P_r\{在传输期之前上一个微时隙有一个数据包到达/源到达\}\\&=\frac{P_r\{一个微时隙内有一个数据包到达\}}{P_r\{一个微时隙内有一些数据包到达\}}\\&=\frac{g\tau e^{-g\tau}}{1-e^{-g\tau}}\end{aligned} \tag{3-62}$$

$$S=\frac{U}{B+I}=\frac{\dfrac{T}{e^{-gt}}\dfrac{g\tau e^{-gt}}{1-e^{-g\tau}}}{\dfrac{T+\tau}{e^{-g\tau}}+\dfrac{\tau}{1-e^{-g\tau}}}=\frac{Tg\tau e}{T+\tau-Te^{-gt}} \tag{3-63}$$

$$S=\frac{aGe^{-aG}}{1+a-e^{-aG}} \tag{3-64}$$

考虑到异步的情况，最后可以检查

$$S_{a\to 0}=\frac{G}{1+G}(未分时隙情况) \tag{3-65}$$

3.3.2.2 **分时隙1-坚持 CSMA 的吞吐量**

对于1-坚持 CSMA 的情况，

$$I=\frac{\tau}{1-e^{-g\tau}} \tag{3-66}$$

$$P_r\{\tilde{B}=k(T+\tau)\}=[1-e^{-g(T+\tau)}]^{k-1}e^{-g(T+\tau)},k=1,2,\cdots \tag{3-67}$$

由于如果在前 $k-1$ 个传输期的上一个微时隙内都至少有一个数据包到达，并且在第 k 个传输期的上一个微时隙内没有数据包到达，繁忙期包括 k 个传输期：

$$B=\frac{T+\tau}{e^{-g(T+\tau)}} \tag{3-68}$$

在繁忙期的第一个传输期成功传输的概率 $P_{\mathrm{SUCC},1}$ 不同于在繁忙期其他传输期成功传输的概率 $P_{\mathrm{SUCC},2}$：

$$\begin{aligned}P_{\mathrm{SUCC},1}&=P_r\{\text{在繁忙期的第一个传输期成功传输}\}\\&=P_r\left\{=\frac{g\tau e^{-g\tau}}{1-e^{-g\tau}}\text{在上一个微时隙到达单个数据包}\mid\text{至少}\right.\\&\qquad\left.\text{有一个数据包到达}\right\}\end{aligned} \tag{3-69}$$

$$\begin{aligned}P_{\mathrm{SUCC},2}&=P_r\{\text{在繁忙期的非第一个传输期成功传输}\}\\&=P_r\left\{=\frac{g\ (T+\tau)\ e^{-g(T+\tau)}}{1-e^{(T+\tau)}}\text{在上一个微时隙到达}\right.\\&\qquad\left.\text{单个数据包}\mid\text{至少有一个数据包到达}\right\}\end{aligned} \tag{3-70}$$

因此，可以得到

$$U=T[P_{\mathrm{SUCC},1}+\left(\frac{B}{T+\tau}-1\right)P_{\mathrm{SUCC},2}] \tag{3-71}$$

$$S=\frac{U}{B+I}=\frac{Ge^{-(1+a)G}(1+a-e^{-aG})}{(1+a)(1+e^{-aG})+ae^{-a(1+a)G}} \tag{3-72}$$

$$S_{a\to 0}=\frac{Ge^{-G}(1+G)}{G+e^{G}} \tag{3-73}$$

前面比较了几种 CSMA 系列协议的性能，从图3-17可以看出分时隙 CSMA 系统可以提供较好的性能，虽然在实际上可以忽略这种性能上的增益。当然，分时隙 CSMA 系统更容易分析。

3.3.3 带碰撞检测的载波监听多址接入（CSMA/CD）

为了改进 CSMA 系统性能，必须减少周期长度。1-坚持协议可以缩短空闲期，但是在大多数情况下1-坚持 CSMA 性能较差。因此，缩短不成功传输时间是惟一改进系统性能的途径。除了检测载波之外，用户可以在一些传输时检测干

扰，并且可以在数据发生碰撞时中止传输。如果这个过程足够快，不成功传输的时间将短于成功传输的时间。这种 CSMA 协议称为带碰撞检测的载波监听多址接入（CSMA/CD）（见图 3-18）。

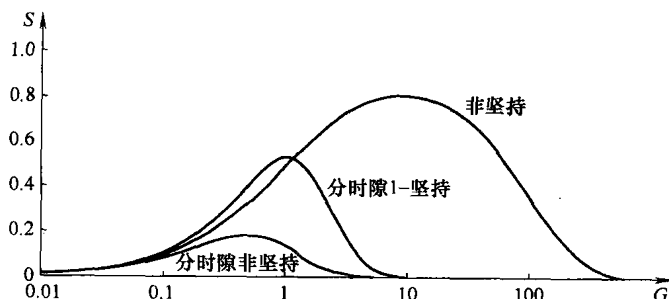


图 3-17 多址接入协议的性能比较

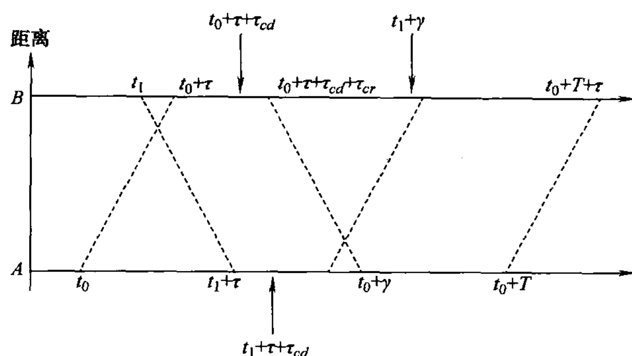


图 3-18 CSMA/CD 示意图

假设用户 A 在 t_0时刻开始传输，此时信道空闲，传输在 $t_0+\tau$ 时到达用户 B。假设 B 在时刻 $t_1 < t_0+\tau$ 开始新的传输（此时 B 监测信道空闲），由于需要 τ_{cd}来检测碰撞，用户 B 在 $t_0+\tau+\tau_{cd}$时刻检测到了碰撞。在很多局域网内一旦检测到发生碰撞，例如以太网，每个用户共同开始一个增强过程，通过一个持续时间 τ_{cr}的碰撞信号来干扰信道，从而确保所有的网络用户都确定发生了碰撞。在 $t_0+\tau+\tau_{cd}+\tau_{cr}$时刻，用户 B 完成上面的增强过程，在 $t_0+2\tau+\tau_{cd}+\tau_{cr}$时刻到达用户 A。从用户 A 来看，这个传输期的持续时间为

$$\gamma = 2\tau + 2\tau_{cd} + \tau_{cr} \tag{3-74}$$

类似的，B 在 $t_1+\gamma$ 时刻完成了该传输期。因此该信道在 $t_1+\gamma-t_0$时间内是繁忙的。在最坏的情况时，在一个不成功的传输期内，信道在时间 $\gamma+\tau$ 内保持繁忙，用 $\widetilde{X}$ 表示传输长度：

$$\widetilde{X} = \begin{cases} T+\tau, \text{成功传输期} \\ \gamma+\tau, \text{不成功传输期} \end{cases} \tag{3-75}$$

3.3.3.1 分时隙非坚持 CSMA/CD 吞吐量

一个周期由一个繁忙期（包括成功传输期和不成功传输期）和一个空闲期组成。空闲期的分布类似于在分时隙非坚持 CSMA 中计算的结论：

$$P\{\tilde{I}=k\tau\}=(e^{-g\tau})^{k-1}(1-e^{-g\tau}),k=1,2,\cdots \tag{3-76}$$

$$I=\frac{\tau}{1-e^{-g\tau}} \tag{3-77}$$

假设 γ 和 T 是 τ 的整数倍，

$$P_{\text{SUCC}}=P_r\{\text{单次传输}|\text{至少一次传输}\}=\frac{g\tau e^{-g\tau}}{1-e^{-gt}} \tag{3-78}$$

每个包含成功传输的传输期长度为 $T+\tau$ 秒，而包含不成功传输的传输期长度为 $\gamma+\tau$ 秒。如果在前 l 传输期的上一个微时隙内至少有一个数据包到达，并且在第 l 个传输期的上一个微时隙内没有数据到达时，繁忙期将包含 l 个传输期。因此，繁忙期正好包括 l 个传输期（$l>1$）的概率为 $e^{-g\tau}(1-e^{-g\tau})^{l-1}$，繁忙期内传输期的平均数量为 $1/e^{-g\tau}$。繁忙期长度的分布概率为

$$\begin{aligned}P_r\{\tilde{B}&=k(T+\tau)+(l-k)(\gamma+\tau)\}\\&=e^{-g\tau}(1-e^{-g\tau})^{l-1}\binom{l}{k}P_{\text{SUCC}}^k(1-P_{\text{SUCC}})^{l-k},l=1,2,\cdots,\ k=0,1,\cdots l\end{aligned} \tag{3-79}$$

当 l 对应繁忙期内传输期总的数量，而 k 对应成功传输期的数量时：

$$\begin{aligned}B&=\sum_{l=1}^{\infty}\sum_{k=0}^{l}[k(T+\tau)+(l-k)(\gamma+\tau)]P\{\tilde{B}=k(T+\tau)+(l-k)(\gamma+\tau)\}\\&=\frac{[P_{\text{SUCC}}(T+\tau)+(1-P_{\text{SUCC}})(\gamma+\tau)]}{e^{-g\tau}}\end{aligned} \tag{3-80}$$

$$\begin{aligned}P_r\{\hat{U}&=kT\}\\&=P_r\{\text{在一个繁忙期内有 }k\text{ 个成功传输期}\}\\&=\sum_{l=k}^{\infty}P_r\{\tilde{B}=k(T+\tau)+(l-k)(\gamma+\tau)\}\end{aligned} \tag{3-81}$$

$$U=\sum_{k=0}^{\infty}kTP_r\{\tilde{U}=kT\}=\frac{T}{e^{-g\tau}}P_{\text{SUCC}} \tag{3-82}$$

归一化吞吐量为

$$S=\frac{U}{B+I}=\frac{aGe^{-aG}}{aGe^{-aG}+(1-e^{-aG}-aGe^{-aG})\gamma'+a} \tag{3-83}$$

式中，$\gamma'=\gamma/T$，当 $\gamma'=1$ 时，S 等效于分时隙非坚持 CSMA。

3.3.3.2 分时隙 1-坚持 CSMA/CD 的吞吐量

注意到，除了第一个传输期的传输成功是由之前的微时隙内到达的数据包

决定，繁忙期内的传输期是否传输成功只取决于之前的传输期长度。如果$\hat{X}_i$表示繁忙期内的第 i 个传输期长度，那么第 $i+1$ 个传输期的长度取决于$\hat{X}_i$，因为第 i 个传输期的类型（成功或碰撞）是由在之前的传输期内到达的数据包数量决定的，而这又是由之前的传输期长度决定的。因此，给定传输期长度为 x，繁忙期剩余的长度是 x 的函数，其平均长度用 $B(x)$ 表示。类似的，信道在繁忙期剩余时间内成功传输的平均时间表示为 $U(x)$。令 $a_i(x)$ 表示在 x 时间内有 i 个数据包到达的概率。由于业务到达服从泊松分布的假设，$a_i(x)=(gx)^i e^{-gx}/i!$。

$$B(x)=\frac{a_1(x)}{1-a_0(x)}[T+\tau+[1-a_0(T+\tau)]B(T+\tau)](\text{成功})$$
$$+\left[1-\frac{a_1(x)}{1-a_0(x)}\right][\gamma+\tau+[1-a_0(\gamma+\tau)]B(\gamma+\tau)](\text{不成功}) \tag{3-84}$$

令 $x=T+\tau$，$\gamma+\tau$，可以得到两个可以很容易求解的等式。$B(x)|_{x=\tau}$可以得到繁忙期的平均长度。类似的，

$$U(x)=\frac{a_1(x)}{1-a_0(x)}[T+[1-a_0(T+\tau)]U(T+\tau)]+$$
$$\left[1-\frac{a_1(x)}{1-a_0(x)}\right][[1-a_0(\gamma+\tau)]U(\gamma+\tau)] \tag{3-85}$$

空闲期的平均长度为 $\tau/(1-e^{-g\tau})$：

$$S=\frac{U}{B+I}=\frac{U(\tau)}{B(\tau)+\tau/(1-e^{-g\tau})} \tag{3-86}$$

虽然上式不是闭合式，计算的性能结果如图 3-19 所示。

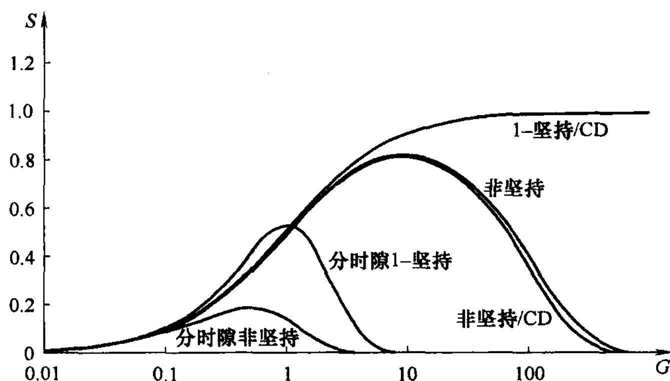


图 3-19 带 CD 的 CSMA 协议比较

请注意，当使用了碰撞检测时，非坚持和1-坚持CSMA之间的性能差距变小；在负荷较小时，1-坚持CSMA/CD的性能很好，因此被广泛使用在局域网中。

3.4 路由

路由算法一般指的是指引数据包通过通信网络到达正确的目的地的网络层协议。在数据网中，同一个用户对两个连续的数据包可能会通过不同的路由传输，对每个单独的数据包都需要做出路由判决。虚拟电路网中，在每条虚拟电路建立时做出路由判决。

路由算法执行的两个主要功能是为不同的源-目的对选择合适的路由，一旦选择好路由后把消息分发到正确的目的地。因此路由算法影响的两个主要性能指标是吞吐量和平均分组延迟。

因为路由算法成功地降低了分组时延，流控制算法允许更多的业务流进入网络（见图3-20）。

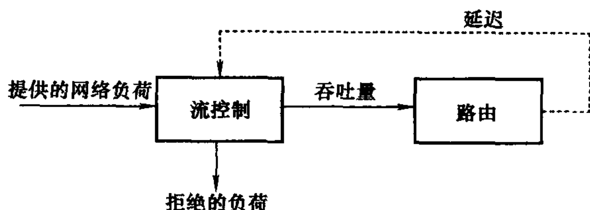


图3-20 路由和流控制的关系

下面将简要地介绍广域网（WAN）的路由方法，可以按以下两种方法分类。如果考虑的是确定性路由方法，我们可以把路由算法分为集中式或分布式。如果考虑的是路由路径，我们可以把路由算法分为静态路由和自适应路由，其中自适应路由是根据最近更新的网络信息搜寻最佳的路径。

3.4.1 洪泛和广播

广播是把数据包从单个发射机传输到单个接收机或一组接收机最简单的方法。虽然效率不高，但是广播方式很合乎情理，因为很简单，并且适合于网络中接收机位置未知的情况。洪泛则是按下面的方法运行。源节点把消息用数据包的方式发送到邻节点。邻节点把数据包中继发射到邻近的节点直到数据包到达网络中所有的节点。通过以下两个附加规则限制了分组传输的数量：

1）节点不会把数据包中继传回已经收到数据包的节点。

2）节点最多只会传输数据包到邻近节点一次。

源节点新产生的数据包中包括源节点 ID 和递增的序号信息，中间节点存储接收到的每个源节点最大的序号的数据包，并且不传输那些序号小于或等于存储序号的数据包，这样可以保证以上两项规则。

另一种广播的方法是使用生成树。生成树是包含所有节点并且没有形成环的网络连接图。基于生成树的广播比洪泛方式效率更高。每个数据包只需要 $N-1$ 次传输，其中 N 是节点数。

3.4.2 最短路径路由

更实际的路由算法是根据两个节点之间的最短路径。每条链路都有一个正数，称为链路长度。一条链路在不同的方向有着不同的长度。两个节点之间的每条路径长度等于该路径上所有链路的长度之和。最短路径路由算法是把每个数据包沿着源节点和目的节点之间长度最短的路径传输。一般的，一条链路的长度取决于链路的传输容量和业务负荷。最短路径应该包括相对小和空闲的链路，因此更适合路由算法。

更复杂的路由算法是随着时间而变化，并且根据链路当前的拥塞程度改变每条链路的长度。最短路径可以随着负荷而变化，并且数据包路由绕开拥塞节点，但是可能会带来路径振荡。

Bellman-Ford 方法是计算到目的节点最短路径的一个重要的分布式算法，其形式是

$$D_i = \min_j [d_{ij} + D_j] \tag{3-87}$$

式中，D_i是节点 i 到目的节点估计的最短距离，d_{ij}是链路（i，j）的长度。实际中，Bellman-Ford 方法可以用迭代的方法实现。

3.4.3 最优路由

最短路径路由有两个缺点：

1）每个源-目的节点对只使用一条路径，因此限制了网络的吞吐量。

2）随着业务状况的改变，路由的能力受限于其对路由振荡的抵抗性。

最优路径根据对平均延迟的最优化，可以通过分解任何源-目的节点对在策略点的业务和在不同路径之间逐渐改变业务来优化测量性能。

3.4.4 热土豆（反射）路由

在每个节点存储空间有限的网络中，更需要修改路由算法以减少缓存溢出和数据包的伴随损耗。其基本思路是节点尽可能快地处理存储的数据包，只要有空闲的链路就立即传输，而不一定是把数据包传到靠近目的地的链路。

3.4.5 切割路由

切割路由的基本思想是利用管道的优点，把长的消息分割成几个小的数据包，从而减少消息在多条链路路径上的延迟。切割路由中，节点可以把数据包的任何一部分中继给其他节点，而不用等待接收到整个数据包。对于一条由 n 条链路组成的路径，其延迟可以减少为原来的 $1/n$。端到端需要使用差错检测重传。

3.4.6 互连网络路由

随着网络规模的扩大，网络之间需要使用不同的接口设备互连。对于广域网，互连的接口称为网关，一般执行相对复杂的网络层任务，例如路由。局域网在 MAC 层用于互连的设备执行简单的路由，称为网桥。局域网之间可以相互连接，也可以连接到广域网，此时使用的互连设备更加复杂，称为路由器（见图 3-21）。

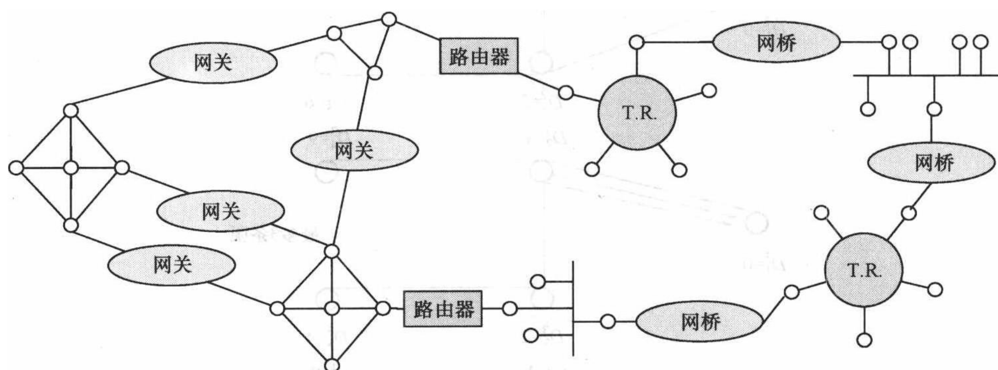


图 3-21 互连网络路由示意图

3.4.7 最短路径路由算法

下面着重介绍最短路径路由算法，因为最短路径算法在认知无线电网络中得到了广泛的应用。

3.4.7.1 Bellman-Ford 算法

假设节点 1 是目的节点，考虑从每个节点寻找到节点 1 的最短路径问题。假设从每个节点到目的节点都存在至少一条路径。如果 (i, j) 不是连接图的弧线，则 $d_{ij} = \infty$。如果限制一条路径最多只包括 h 条弧线，并且只能通过节点 1 一次，从节点 i 到节点 1 的最短通路用 $D_i^h D_i^h = 0$，$\forall h$ 表示（见图 3-22）。

D_i^h可以用下列迭代计算

$$D_i^{h+1} = \min_j [d_{ij} + D_j^h], \forall i \neq 1 \tag{3-88}$$

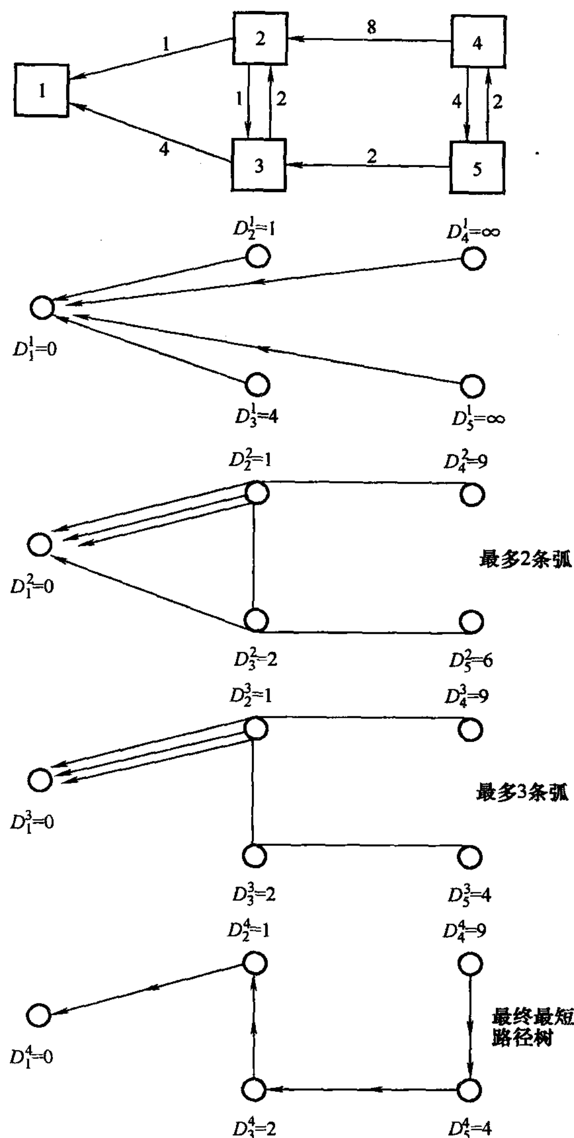


图 3-22 Bellman-Ford 算法的示意图

从初始状态开始

$$D_i^0=\infty\ ,\ \forall i\neq 1 \tag{3-89}$$

我们认为 Bellman-Ford 算法首先寻找单弧最短通路长度，然后计算二弧最短通路，依此类推。我们还认为，如果假设所有的不包括节点 1 的圈长度非负，那么最短通路长度等于最短路径长度。如果满足

$$D_i^h = D_i^{h-1}, \forall i \tag{3-90}$$

该算法将在 h 次迭代之后终止。

命题3.1：考虑初始状态为 $D_i^0 = \infty$，$\forall i \neq 1$ 的 Bellman-Ford 算法，那么：

1）算法计算的 D_i^h 等于从节点 i 到节点1的最短（$\leqslant h$）通路长度。

2）当且仅当所有的不包括节点1的圈长度非负，算法在有限次迭代后终止。此外，如果算法终止，那么算法最短迭代了至多 $h \leqslant N$ 次，在终止后，D_i^h 是从节点 i 到节点1的最短路径长度。

3.4.7.2 Dijkstra 算法

该算法需要所有的弧长非负。在最差情况下，Dijkstra 算法所需的计算复杂度比 Bellman-Ford 算法低（见图3-23的比较）。Dijkstra 算法的基本思路是按路径长度递增的顺序寻找最短路径。到节点1的最短路径中，最短的一定是从节点1最近的临近节点的单弧路径，因为假设弧长非负，所以任何多弧路径都不会比单弧路径长度短。最短路径中，次最短的路径也一定是从节点1的次临近节点的单弧路径或者是通过之前选择的最近节点的最短两弧路径，依此类推。

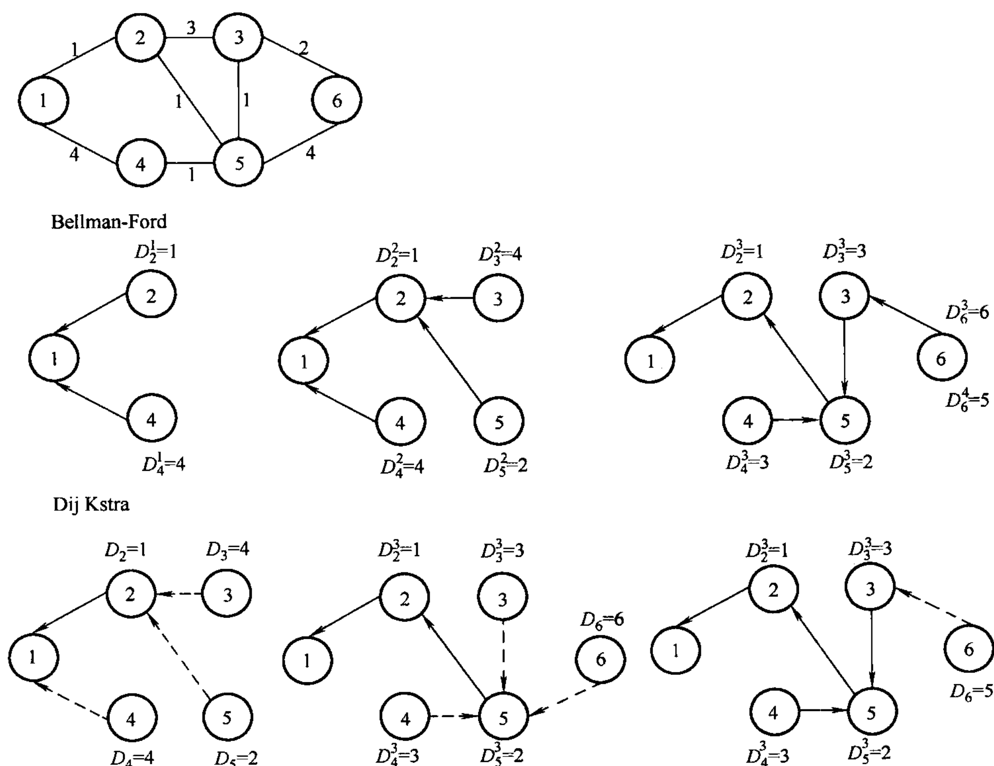


图3-23 Bellman-Ford 算法和 Dijkstra 算法

我们把每个节点 i 到节点 1 的最短路径长度估计标记为 D_i。当估计的最短路径长度确定时，我们把该节点永久标记，并且跟踪永久标记节点集。每一步增加到 P 的节点都是那些不在 P 内节点中到节点 1 最近的节点。具体算法如下：

初始化，$P=\{1\}$，$D_1=0$，$D_j=d_{j1}$，$\forall j\neq 1$

步骤 1：（找到下一个最近节点）找到 $i\notin P$

$$D_i=\min_{j\notin P}D_j \tag{3-91}$$

集合 P：$PU\{i\}$。如果 P 包含所有节点，则停止。

步骤 2：（更新标记）对所有 $j\notin P$

$$D_j:=\min[D_j,d_{ij}+D_i] \tag{3-92}$$

返回步骤 1。

在每个步骤 1 的开始时：

a）$D_i\leqslant D_j$，$\forall i\in P$，$j\notin P$。

b）对每个节点 j，D_j是从 j 到节点 1 的除了 j（属于集合 P）之外所有节点的最短路径。

3.4.7.3 **最优路由**

单调递增函数（代价函数）是

$$\sum_{(i,j)}D_{ij}(F_{ij})$$

经常使用的函数式为

$$D_{ij}(F_{ij})=\frac{F_{ij}}{C_{ij}-F_{ij}}+d_{ij}F_{ij} \tag{3-93}$$

式中，C_{ij}是链路 (i,j) 的传输容量，用与 F_{ij}一样的单位来度量。F_{ij}是链路 (i,j) 的总流量

$$F_{ij}=\sum_{\text{所有包含}(i,j)\text{的路径}p}x_p \tag{3-94}$$

式中，x_p是路径 p 的流。对每个源-目的对 w，存在限制

$$\sum_{p\in P_w}x_p=r_w,x_p\geqslant 0,\forall p\in P_w \tag{3-95}$$

式中，r_w是源-目的对给定的业务输入，P_w是 w 的有向路径集合。对于不确定路径流矢量 $x=\{x_p|p\in P_w,\ w\in W\}$，问题是

$$\text{最小化}\sum_{(i,j)}D_{ij}\Big[\sum_{\text{所有包含}(i,j)\text{的路径}p}x_p\Big]$$

$$\text{受限于}\sum_{p\in P_w}x_p=r_w,\forall w\in W$$

$$x_p\geqslant 0,\forall p\in P_w,w\in W \tag{3-96}$$

假设每个 D_{ij}是 F_{ij}的可微函数，并且定义取值在 $[0,C_{ij}]$ 内，其中 C_{ij}是一个正数或正无穷。令 x 表示路径流 x_p的矢量

$$D(x) = \sum_{(i,j)} D_{ij}[\sum_{\text{所有包含}(i,j)\text{的路径}p} x_p] \tag{3-97}$$

那么，

$$\frac{\partial D(x)}{\partial x_p} = \sum_{\text{路径}p\text{的所有链路}(i,j)} D'_{ij} \tag{3-98}$$

式中，一阶导数 D'_{ij}是对应 x 的所有数据流。$\partial D(x)/\partial x_p$是当每条链路（$i$，$j$）的长度等于在 x 处得到的一阶导数 D'_{ij}时路径 p 的长度。因此，在下文中，称 $\partial D(x)/\partial x_p$为路径 p 的一阶导数长度。

令 $x^* = \{x_p^*\}$表示最优路径流矢量。那么，如果对某个源-目的节点对 w 的某条路径 p，有 $x_p^* > 0$，我们必然能从路径 p 移动一个小的量 $\delta > 0$ 到同样的节点对的其他路径 p'，并且不会增加代价，否则就违背了最优的假设。为了得到 x_p^*，

$$\delta \frac{\partial D(x^*)}{\partial x'_p} - \delta \frac{\partial D(x^*)}{\partial x_p} \geqslant 0 \tag{3-99}$$

$$\frac{\partial D(x^*)}{\partial x'_p} \geqslant \frac{\partial D(x^*)}{\partial x_p}, \forall p' \in P_W \tag{3-100}$$

最优路径流只在一阶导长度最小的路径上为正数。此外，在最优路径流，沿着源-目的节点对 w 分割的输入流 r_w必须有同样的长度（比 w 所有其他路径要短或者等长度）。如果 D_{ij}是凸弧，上面的条件是最优路径流的充分必要条件。

在一些数据网或者在源节点不知道详细路径描述的网络中，可能需要在每个节点 i 保持一个路由表，包括到目的节点 j 每条链路（i，j）的路由变量 $\phi_{ik}(j)$。路由变量是所有到达节点 i、目的节点为 j，并且沿着链路（i，j）的流的部分比例：

$$\phi(k,j) = \frac{f_{ik}(j)}{\sum_m f_{im}(j)}, \forall (i,k), j \tag{3-101}$$

式中，$f_{ik}(j)$ 是沿着链路（i，j）并且目的节点为 j 的流。如果给定了用路径流矢量$\{x_p^*\}$表示的路由问题最优解决方案，那么可能确定$f_{ik}(j)$ 和 $\phi_{ik}(j)$。

3.5 流控制

在大多数网络中，有可能遇到即使使用最优路径，外部提供的带宽仍然比能处理的带宽要大的情况。那么，如果不采取限制业务输入网络的措施，那么网络瓶颈链路的数据队列长度会持续增长，分组延迟也会增大，从而出现类似高速公路上的交通阻塞。

一般当由于通信线路容量有限或者处理硬件能力有限，需要限制两个节点的通信速率时，需要采用流控制。流控制可能用在传输层、网络层或因特网层。

流控制的方法包括以下几种：

1）呼叫阻塞；

2）丢弃数据包；

3）阻塞数据包；

4）调度数据包。

流控制的主要目的如下：

1）在节流会话（满足最小数据速率需求）、保持合适的平均时延和缓存之间折中。

2）保证会话间服务质量的公平性。

3.5.1 窗口流控制

如果接收机A发送数据单元的数量存在上限，并且A不知道B接收到的数据单元数，发射机A和接收机B之间的会话称为窗口流控制，该上限（正整数）称为窗口大小。

接收机B处理完数据单元之后会发送一个特别的消息通知发射机A，该消息称为许可。一旦接收到许可，A可以再发送一个数据单元给B。窗口策略的基本思路是当许可返回很慢时，降低发射机的输入速率。因此，如果在会话中的通信路径中发生拥塞时，许可较大的延迟会造成发射机的数据速率下降。然而，窗口策略还有其他用途，例如接收机可以通过延迟许可来限制会话的传输速率，从而避免缓存溢出。

窗口流控制一般是在端到端执行。在大多数情况下，窗口大小为αW，其中α和W是正数。当目的节点每次接收到新的α个数据单元时，目的节点将返回一个许可到源节点，从而分配新的一批α个数据单元。

一般可使用一种分组和许可的编号方案把许可与之前传输的数据包联系在一起，丢失的许可可以使用类似数据链路控制中使用的协议，其中数据包包括一个序号和一个请求号。请求号可以在一个或多个许可中用于流控制。例如，假设节点A从节点B接收到一个请求号为k的数据包。A知道B已经处理完A发送的所有编号小于k的数据包，A可以发送最多编号为$k+W-1$的未发送数据包（W为窗口大小）。在这种方案中，一旦在发射机和接收机之间建立起数据包的顺序，序号和请求号都用模m（$\geq W+1$）的方式表示。

假设源节点只计算已经发送但是没有收到返回的许可的数据包数量x，并且只传输新的长度为$x<W$的数据包。

考虑往返时延d包括往返传播时延的数据包流情况，数据包传输时间和许可延迟比传输整个窗口W个数据包所需的时间要短，也就是

$$d \leqslant WX \tag{3-102}$$

式中，X 是一个数据包的传输时间。那么源节点可以全速率 $1/X$ 传输，并不使用流控制。

对于 $d > WX$ 且往返时延 d 足够大，完整的 W 个数据包可以在第一个许可返回之前传输完。假设源节点一直有数据包在队列中等待传输，那么传输速率为 W/d。

对应往返时延 d 的最大传输速率为

$$r = \min\left\{\frac{1}{x}, \frac{W}{d}\right\} \tag{3-103}$$

还要注意，端到端的窗口不能保证最小通信速率（见图3-24）。与此同时，我们一般需要优化无线网络的窗口大小。

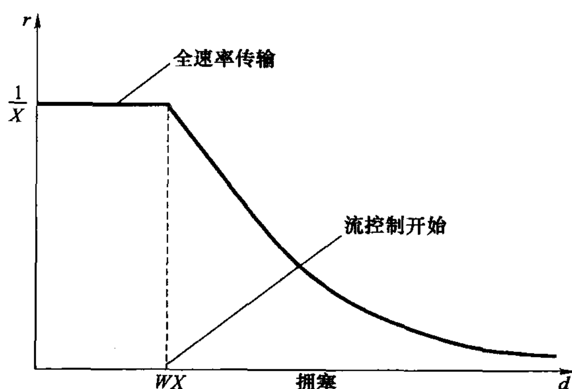


图3-24 端到端窗口

3.5.2 速率控制方案

窗口流控制并不适合于高速广域网的高速会话。另一种流控制的方法是根据每个会话的需求保证每个会话的数据速率。该速率应该根据会话类型有一定的限制。确定输入会话速率的主要考虑因素包括：

1）延迟-吞吐量折中；

2）公平性：如果必须减少会话速率以容纳新的会话，那么在保证每个会话的最小速率需求的同时，必须公平的降低速率。

更合适的办法是考虑突发业务，每隔 W/r 秒允许 W 个数据包接入网络。对每个会话分配 W 个数据包，在会话开始时，保持分配的未使用部分的数量为 x。只要 $x>0$，那么将允许会话的数据包都接入网络。每次当有一个数据包被允许传输时，该数量 x 减1。这种流控制类似于窗口长度为 W 的窗口流控制，除了在速率流控制中，该数量在接纳每个数据包后每隔 W/r 秒加1，而在窗口流控制中

该数量是当对应的许可返回时在往返时延后加 1。

相应的管理传输业务突发性更好的方法是漏桶方案。在这种方案中数据包数量是周期性的每隔 $1/r$ 秒加 1，最大可达到 W 个数据包。从另一个角度来看，对每个会话，有一队数据包没有许可，在会话的源节点有一桶许可。拥有许可的队列头的数据包等待被传输。许可按会话需要的输入速率 r 产生（每隔 $1/r$ 秒产生一个许可），只要许可桶中的许可数量不超过一定的门限 W（见图 3-25）。

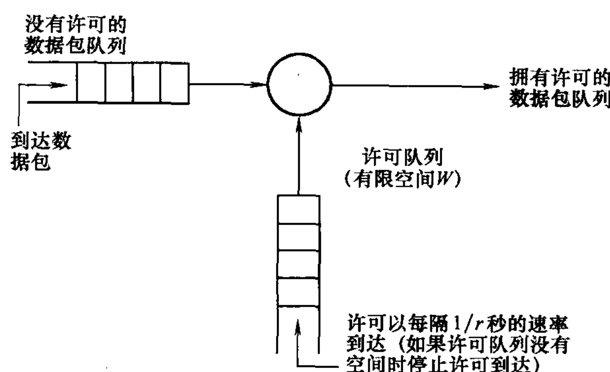


图 3-25 数据包获得许可的过程

漏桶方案不一定能完全阻止缓存溢出，并且如果没有其他选择和执行网络内会话速率的机制，也不能确保分组时延上限。

通过选择合适的桶大小参数，可以减少缓存溢出和最大分组时延。实际上，桶的大小 W 是影响漏桶方案性能最重要的参数之一。如果 W 很小，突发业务被延迟等待传输许可（$W=1$ 等效于时分复用）。如果 W 过大，长的突发数据包将被允许进入网络，这些突发数据包会在下行拥挤节点处堆积，引起缓存溢出。因此有人提出了动态调整桶大小的想法，它与动态调整许可速率可以合并到一个算法。然而，在高速网络中，拥塞节点传输的反馈控制消息可能会由于相对大的传输时延而失效。因此，可能需要某些预测机制在拥塞发生前发送控制消息。

3.5.3 漏桶方案的队列分析

在分析之前，我们做了如下假设：

1）数据包到达服从速率 λ 的泊松过程。

2）许可每隔 $1/r$ 秒到达，但是如果许可池包括 W 个许可，新到达的许可将会被丢弃。

我们把这个系统看成一个离散时间马尔科夫链，其状态为 0，1，…，状态 $i=W+1$，$W+2$，…，对应 $1-W$ 个没有可用许可、处于等待中的数据包。状态

$i=0$，1，…，W 对应有 $W-i$ 个可用许可、未等待的数据包。状态转移在许可到达后0，$1/r$，$2/r$，…时发生。在 $1/r$ 秒内有 k 个数据包到达的概率为

$$a_k=\frac{e^{-\frac{\lambda}{r}}\left(\frac{\lambda}{r}\right)^k}{k!} \tag{3-104}$$

可以看出，马尔科夫链的转移概率为

$$P_{0i}=\begin{cases}a_{i+1},i\geqslant 1\\ a_0+a_1,i=0\end{cases} \tag{3-105}$$

对于 $j\geqslant 1$

$$P_{ji}=\begin{cases}a_{i-j+1},j\leqslant i+1\\ 0,\text{其他}\end{cases} \tag{3-106}$$

可以得到全局平衡式

$$P_0=a_0P_1+(a_0+a_1)P_0 \tag{3-107}$$

$$P_i=\sum_{j=0}^{i+1}a_{i-j+1}P_j,i\geqslant 1 \tag{3-108}$$

可以递归求解这些等式，并可以得到状态稳定概率为

$$P_0=\frac{r-\lambda}{ra_0} \tag{3-109}$$

注意，当数据包到达速率为 λ 时，所有状态平均的许可生成速率为 $(1-P_0a_0)r$。如果 $\lambda<r$，则系统是稳定的，也就是数据包队列有界。一个数据包获得一个许可的平均时延为

$$\begin{aligned}T&=\frac{1}{r}\sum_{j=0}^{\infty}P_j\max\{0,j-w\}\\&=\frac{1}{r}\sum_{j=w+1}^{\infty}P_j(j-w)\end{aligned} \tag{3-110}$$

为了得到数据包获得许可所需的平均时延的闭合表达式，也为了得到实际系统更好的模型，我们稍微修改了漏桶方案模型，让许可按每个比特产生。这个假设更符合实际情况，因为实际上消息都在到达源节点后分解成小的数据包。我们还做了如下特别的假设：

1）允许进入网络的许可以 r 比特/秒的速率生成，漏桶的大小为 w 比特。

2）消息到达服从速率为 λ 的泊松过程，消息的存储空间无限。消息长度相互独立，并且服从均值为 L 比特的指数分布。

令 $\mu=r/L$，因此许可消息用速率 r 传输所需的平均时间为 $1/\mu$。令 $c=w/r$ 表示传输许可可以节省的时间。系统的状态可以用队列中的比特数和可用许可数量来描述。在时间 t，令 $X(t)$ 表示队列中的比特数（如果队列非空）或负的可用许可数。

因此，当一个 x 比特的消息到达时，$X(t)$ 增加 x。(下列三件事中的一件将发生：许可减少 x；队列长度增加 x；或者如果 $X(t)<0$，$X(t)+x>0$，许可减少到0并且队列增加到 $X(t)+x$)。

令 $Y(t)=X(t)+c$，可以看出 $Y(t)$ 是到达速率为 λ 消息/秒和服务速度为 μ 的 $M/M/1$ 虚拟队列中未完成的工作。如果虚拟队列前的数据比特小于 c，进入队列的比特将会立即传输，否则，进入队列的比特将会比虚拟队列中提前 L 秒传输。考虑消息的最后比特，我们可以看出如果 T_i 是第 i 个消息队列的系统延迟，$\max\{0, T_i-c\}$ 是实际队列的延迟。虚拟队列系统时间 T_i 的稳定状态分布为

$$P\{T_i \geqslant \tau\} = e^{-\tau(\mu-\lambda)} \tag{3-111}$$

令 $T'_i=\max\{0, T_i-c\}$ 表示第 i 个数据包在实际队列中的延迟，那么

$$P\{T'_i \geqslant \tau\} = \begin{cases} 1, & \tau \leqslant 0 \\ e^{-(c+\tau)(\mu-\lambda)}, & \tau > 0 \end{cases} \tag{3-112}$$

实际队列中数据包的平均时延为

$$T = \int_0^{\infty} P\{T'_i \geqslant \tau\} = \frac{1}{\mu-\lambda} e^{-c(\mu-\lambda)} \tag{3-113}$$

如果读者想深入了解计算机网络中研究的技术，我们推荐查阅 Bertsekas 和 Gallager 的数据网络[1]。然而，作为本书后面章节中 Ad hoc 网络很多概念的基础，本章中的描述已经足够。

参考文献

[1] D. Bertsekas, R. Gallager, *Data Networks*, Prentice-Hall, New Jersey, 1992.

[2] N. Abramson,'The ALOHA System – Another Alternative for Computer Communications', *AFIPS. Conf. Proc.* **37**. Montvale. N.J. (ed.), AFIPS Press, in *1970 Fall Joint Computer Conf.*, 1970, pp. 281–285.

[3] F.A. Tobagi, V.B. Hunt, 'Performance Analysis of Carrier Sense Multiple Access with Collision Detection', *Computer Networks*, **4**,November 1980.

[4] G. Frenkel, 'The Grade of Service in Multiple Access Satellite Communication with Remand Assignment' *IEEE Tr. on Communications (Technology)*, **12**(10), 1972, 1681–1685.

[5] L. Kleinrock, F. Tobagi, 'Packet Switching in Radio Channels, Part I – Carrier Sense Multiple Access and Their Throughput – Delay Characteristics', *IEEE Tr. on Communications*, **23**(12), 1975, 1400–1416.

[6] L. Kleinrock, F. Tobagi. 'Packet Switching in Radio Channels, Part II – The Hidden Terminal Problem and Busy Tone Solution', *IEEE Tr. on Communications*, **23**(12), 1975, 1400–1416.

[7] N. Abramson, 'The Throughput of Packet Broadcasting Channels', *IEEE Tr. on Communications*, **25**(1), 1977, 117–128.

[8] J. Hayes, 'An adaptive Technique for Local Distribution', *IEEE Tr. on Communications*, **26**(8), 1978, 1178–1186.

[9] W. Bux, 'Local Area Subnetworks: A Performance Comparison', *IEEE Tr. on Communications*, **29**(10), 1981.

[10] E. Cole, B. Liu, 'Finite Population CSMA-CD Networks', *IEEE Tr. on Communications*, **31**(11), 1983, 1247–1251.

[11] J. Meditch, C.T. Lea, 'Switching and Optimization of the CSMA and CSMA/CD Channels', *IEEE Tr. on Communications*, **31**(6), 1983, 763–774.

[12] H. Takaqi, 'Queuing Analysis of Polling Models', *ACM Computing Surveys*, **20**,1988, 5–28.

[13] T. Kample, 'Optimal Scheduling of Jobs with Exponential Service Times on Identical Parallel Processors', *Operation Research*, **37**,1988, 126–133.

[14] F. Baccelli, M. Malcowski, 'Queuing Models for Systems with Synchronization Constraint', *IEEE Proc.*, **39**(1), 1989, 138–161.

[15] J. Hui, 'Resource Allocation for Broadband Networks', *IEEE Journal on Selected Areas in Communications*, **6**(9), 1989, 1598–1618.

[16] N. Abramson, 'Fundamental of Packet Multiple Access for Satellite Networks' *IEEE Journal on selected Areas in Communications*, **10**(1), 1992, 309–316.

[17] A. Ephremides, B. Hajek, 'Information Theory and Communications Networks: An Unconsummated Union', *IEEE Transactions on Information Theory*, **44**(6), 1998, 2416–2434.

[18] K.C. Chen, 'Medium Access Control of Wireless LANs for Mobile Computing', *IEEE Networks*, 1994.

第 4 章　协作通信和网络

MIMO 通信已经成功的提高了系统容量和系统在严重衰落中的覆盖，如在第 1 章描述的，MIMO 使用了很多数量的传输天线和分集天线，以提高分集增益。然而，在终端设备上由于设备尺寸有限不能使用很多天线，但是我们可以使用中继节点以扩大系统覆盖。研究者还指出，我们可以使用协作分集来利用分布式单天线节点的空间分集，从而为协作通信和协作网络打开了一扇门。

协作分集的基本思路是把一个信号通过不同天线的独立信道进行传输和中继，利用空间分集增益来实现传输分集，如图 4-1 所示。源节点 A 传输源信号 X 到目的节点 C 和中继节点 B，中继节点 B 传输信号 X_1（包括了 X 的信息）到目的节点 C，这样就构成了最简单的三节点（或称三终端）中继网络。目的节点 C 接收到两个信号，Y 和 Y_1。这两个信号包含了 X 通过不同路径传递的信息，因此引入了空间分集，从而构成了协作通信或协作分集。

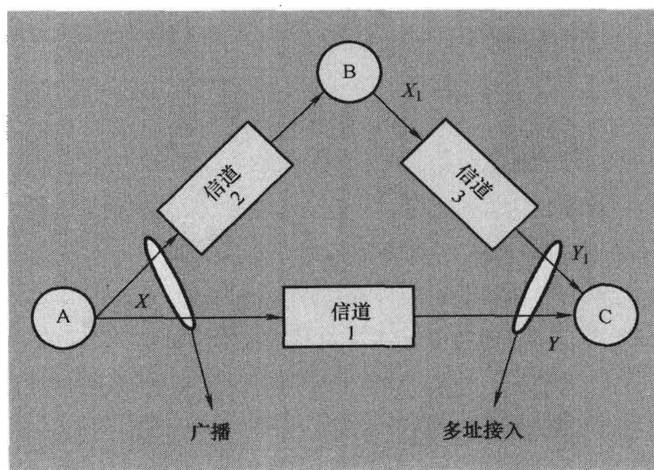


图 4-1　中继信道

此外，除了增加对抗衰落的分集阶数，协作通信和协作网络（因为其中包括三个或三个以上的节点）还存在其他好处，例如通过协作中继增加了累计带宽。

4.1　协作通信的信息论

协作（分集）通信是通过在时域、频域或空域上合并直接传输的原始信号

和中继传输的信号实现的，并且把中继的信号进行译码，而不是把中继信号当做干扰。在本节，我们将介绍信息论和协作通信的基础。Cover 和 El Gamal 在参考文献 [3] 中研究了中继信道的信道容量，其中考虑了包括源节点、中继节点和目的节点的三节点/终端中继网络。该研究结果表明协作通信的容量比非协作通信的容量大。

我们将从简单的点对点信号的容量区域开始，然后介绍了多址接入信道（上行信道）、广播信道（下行信道）、中继信道和干扰信道的容量，最后在介绍协作协议之前，我们分析了具有协作分集的多址信道容量。

4.1.1　基本网络信息论

令 X 表示信道的随机输入，Y 表示信道的随机输出。最简单的点对点链路的容量为 $C = \max_{p(x)} \mathrm{I}(X;Y) = \log\left(1+\frac{P}{N}\right)$，其中 P 是发射功率限制，N 是接收机加性白噪声（AWGN）功率。

图 4-2 给出了最简单的两个发射机的多址信道示例。

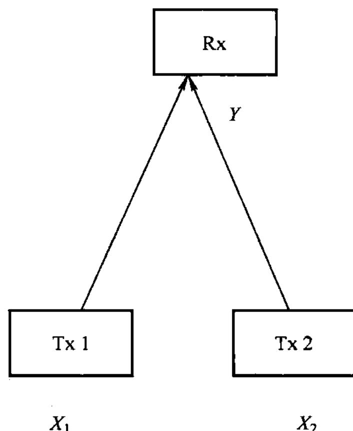


图 4-2　简单的多址接入信道

两个发射机 Tx1 和 Tx2 与一个接收机通信。接收的信号 $Y = X_1 + X_2 + Z$，其中 Z 是方差为 N_0 的循环对称复高斯噪声。发射机 j 有功率限制 $P_j(j=1,\ 2)$。众所周知，容量区域是满足下列条件的速率对凸包集：

$$R_1 \leqslant \mathrm{I}(X_1;Y|X_2) = \log\left(1+\frac{P_1}{N_0}\right) \tag{4-1}$$

$$R_2 \leqslant \mathrm{I}(X_2;Y|X_1) = \log\left(1+\frac{P_2}{N_0}\right) \tag{4-2}$$

$$R_1 + R_2 \leqslant \mathrm{I}(X_1;X_2;Y) = \log\left(1+\frac{P_1+P_2}{N_0}\right) \tag{4-3}$$

其中，$X_1 \sim N(0,P_1)$，$X_2 \sim N(0,P_2)$时等式成立。

这种最简单的多址接入信道的容量区域如图 4-3 所示。

这种可以到达容量区域拐点的方案可以看成是一个两状态过程。例如，我们考虑 A 处：在第一阶段，接收机译码发射机 Tx2 的信号，并把发射机 Tx1 的信号当做噪声处理。如果 $R_2 \leqslant \mathrm{I}(X_2;\ Y) = \log\left(1+\frac{P_2}{P_1+N_0}\right)$，译码的错误率将较低。

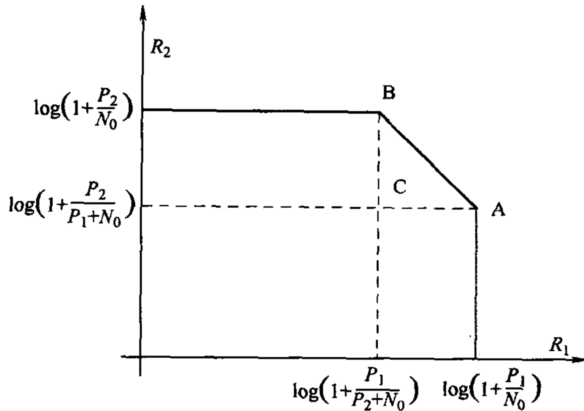


图 4-3 简单的多址接入信道的容量区域

在发射机 Tx2 成功译码后，可以从接收信号中抵消掉 Tx2 的信号，如果 $R_1 \leqslant I(X_1; Y \mid X_2) = \log\left(1+\frac{P_1}{N_0}\right)$，就能准确地译码出 Tx1 发射的信号。这种办法的思路等效于连续干扰抵消。

接下来我们讨论图 4-4 所示的拓扑中广播信道的容量。

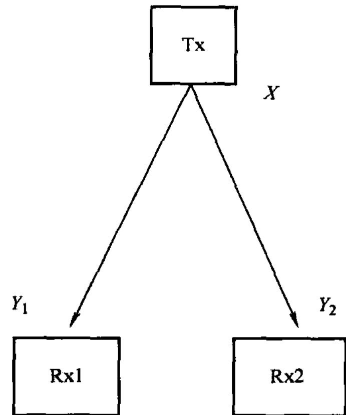


图 4-4 广播信道

有一个功率限制为 P 的发射机和两个接收机 Rx1、Rx2。接收到的信号是 $Y_1 = X + Z_1$、$Y_2 = X + Z_2$。注意 $X = X_1 + X_2$，式中 X_1和 X_2分别是以功率 αP 和 $(1-\alpha)P$ 发送到 Rx1 和 Rx2 的信号。Z_1和 Z_2是方差为 N_1和 N_2的独立循环对称复高斯噪声（假设 $N_1 < N_2$）。发射机希望以速率 R_1和 R_2分别发射独立的消息 X_1和 X_2到接收机 Rx1 和 Rx2。

高斯广播信道的容量区域是

$$R_1 \leqslant \log\left(1+\frac{\alpha P}{N_1}\right) \quad R_2 \leqslant \log\left(1+\frac{(1-\alpha)P}{\alpha P + N_2}\right) \tag{4-4}$$

式中，$0 \leqslant \alpha \leqslant 1$，可以任意选择 α 在速率 R_1和速率 R_2间折中。

能达到容量区域的方案可以按下面的分析考虑。首先，考虑接收机 Rx2 $(N_1 < N_2)$，由于 Rx1 的消息被当做噪声处理，因此信噪比为$\frac{(1-\alpha)P}{\alpha P + N_2}$。因此，Rx2 以速率 $R_2 \leqslant \log\left(1+\frac{(1-\alpha)P}{\alpha P + N_2}\right)$传输。然后，考虑接收机 Rx1，由于噪声功率 N_1较低，所以可以译码 Rx2 的消息。然后在接收信号中抵消 Rx2 的消息，并译

码 Rx1 的消息。因此，Rx1 可以速率 $R_1 \leqslant \log\left(1+\frac{\alpha P}{N_1}\right)$ 传输。

最简单的中继信道由一个源节点、一个目的节点和一个中继节点组成，如图 4-5 所示。

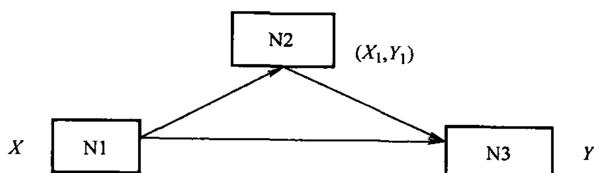


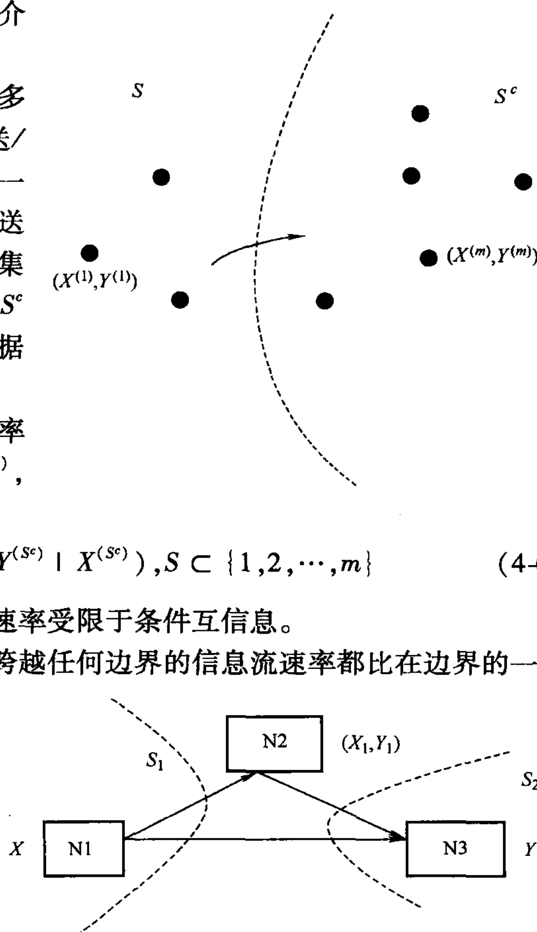
图 4-5　中继信道

N1 是源节点，N3 是目的节点，N2 是中继节点。中继信道合并了一个广播信道（N1 到 N2 和 N3）和一个多址接入信道（N1 和 N2 到 N3）。N2 从 N1 接收了 Y_1，并向 N3 传输 X_1。

中继信道的容量上限是

$$C \leqslant \max_{p(x,x_1)} \min\{I(X,X_1;Y\},I(X;Y,Y_1 \mid X_1)\} \tag{4-5}$$

中继信道的容量可以根据下面介绍的一般最大流最小割理论得到。



考虑一般的由 m 个节点组成的多终端网络，每个节点 i 与一个发送/接收消息对（$X^{(i)}$，$Y^{(i)}$）联系在一起。节点 i 以速率 $R^{(ij)}$ 向节点 j 发送信息。现在，我们把节点分成两个集合，S 和 S 的补集 S^c。S 中节点到 S^c 中节点的信息流速率的上限可以根据下列定理得到：

定理 4.1：如果可以到达信息速率 $\{R^{(ij)}\}$，存在某种联合概率分布 $p(x^{(1)}, x^{(2)}, \cdots, x^{(m)})$，并且对所有满足

$$\sum_{i \in S, j \in S^c} R^{(ij)} \leqslant I(X^{(S)};Y^{(S^c)} \mid X^{(S^c)}), S \subset \{1,2,\cdots,m\} \tag{4-6}$$

因此，跨越割集的信息流的总速率受限于条件互信息。

备注：上述最大流最小割说明跨越任何边界的信息流速率都比在边界的一侧输入、在另一侧输出的互信息（条件是在另一侧输入）要小。还要注意，即使是对于简单的信道，一般也很难达到容量的上限。

让我们用最大流最小割定理来验证中继信道的容量界（见图 4-6）。

图 4-6　中继信道模型的割

$$C \leq \max_{p(x,x_1)} \min\{I(X,X_1;Y), I(X;Y,Y_1|X_1)\} \tag{4-7}$$

式中，第一项 $I(X, X_1; Y)$ 表示从发送方 N1 和 N2 到接收方 N3（割 S_2）的最大信息速率。第二项 $I(X; Y, Y_1 | X_1)$ 表示从 N1 到 N2 和 N3（割 S_1）的最大信息速率。

退化中继信道（也就是 $I(X; Y, Y_1 | X_1) = I(X; Y_1 | X_1)$）的容量是可以达到的。这种思想称为分组马尔科夫编码和译码转发两种方案。

我们把信息编码成 B 个分组，每个分组长度为 n。这 B 个分组使用了同样的长度为 n 的码本。$W_j[i] \in M_j = \{1,2,\cdots,2^{nR_j}\}$ 是发送端 $j(j=1)$ 在第 $i(i=1, 2, \cdots, B)$ 个分组中传输的消息。W 可以编码成 $x_{j'}[i'](w_j[i])$。（还可以表示为 $x_{j'}^n[i'](w_j[i])$；我们使用第一个矢量记号。）$x_{j'}[i'](.)$ 是随机高斯码本 C_j 的一个码组，i' 是分组的下标，j' 是发送端的下标。考虑在节点 1 和 3 之间传输，并且节点 2 作为中继节点的中继信道。P_i 是发送节点 i 的功率（$i=1, 2$）。N_i 是第 i 个节点 AWGN 噪声功率（$i=2, 3$）（见图 4-7）。

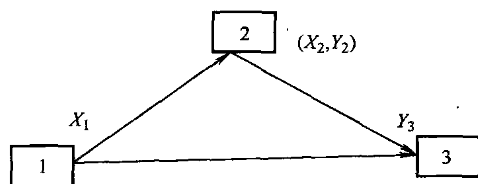


图 4-7　方案 1（译码转发 + 并行信道译码）

中继信道的过程如下：

1）节点 1 传输分组序列 $x_1[1](w_1[1]), \cdots, x_1[B](w_1[B])$。

2）中继节点 2 在第 i 个分组译码 $w_1[i]$，如果 $R_1 \leq \log\left(1+\frac{P_1}{N_2}\right)$，那么可能没有误码产生。然后节点 2 在第 $i+1$ 个分组传输 $w_1[i]$，即节点 2 传输序列 0, $x_2[2](w_1[1]), \cdots, x_2[B](w_1[B])$。这种方法称为译码转发。

3）在接收节点 3，提前进行译码。假设已经接收到分组 i，并且 $w_1[i-2]$ 已经准确译码，因此已知 $x_2[i-1](w_1[i-2])$。根据分组 i 和 $i-1$ 的接收信号：

$$\begin{bmatrix} y_3[i] \\ y'_3[i-1] \end{bmatrix} = \begin{bmatrix} y_3[i] \\ y_3[i-1] - x_2[i-1](w_1[i-2]) \end{bmatrix}$$
$$= \begin{bmatrix} x_1[i](w_1[i]) + x_2[i](w_1[i-1]) + z_3[i]) \\ x_1[i-1](w_1[i-1]) + z_3[i-1] \end{bmatrix}$$

在译码 $w_1[i-1]$ 时，信号 $x_1[i](w_1[i])$ 可以看成是背景噪声的一部分。两个信号 $(y_3[i], y'_3[i-1])$ 组成一个并行高斯信道。速率是

$$R_1 \leq \log\left(1+\frac{P_2}{P_1+N_3}\right) + \log\left(1+\frac{P_1}{N_3}\right) = \log\left(1+\frac{P_1+P_2}{N_3}\right)(=I(X_1,X_2;Y_3)) \tag{4-8}$$

这种方法称为并行信道译码。

方案 2（译码转发 + 后向译码）是从复用编码的概念引申而来。假设发射机

想传输两个消息 $w_1 \in \{1,\cdots,2^{nR1}\}$ 和 $w_2 \in \{1,\cdots,2^{nR2}\}$。发射机生成一个 2^{nR1} 行和 2^{nR2} 列的表，表里的元素是随机分配的高斯码。如果信道容量 $C > R_1 + R_2$，接收机可以译码 w_1 和 w_2。如果接收机已知 w_1，并且 $C > R_2$，那么接收机可以通过根据 w_1 搜索行来译码 w_2，从而得到 w_2。步骤如下：

1）源节点1传输 $x_1[1](0,w_1[1])$，$x_1[2](w_1[1], w_1[2]),\cdots$，$x_1[B+1](w_1[B], 0)$，其中使用复用编码来生成 $x_1[i](w_1[i-1], w_1[i])$。

2）中继节点2仍然译码转发。假设节点2在分组 $i-1$ 译码 $w_1[i-1]$。如果 $R_1 \leqslant \log\left(1+\frac{P_1}{N_2}\right)$，那么节点2可以根据 $x_1[i]$ $(w_1[i-1], w_1[i])$ 译码 $w_1[i]$。

3）目的节点3接收到 $y_3[i] = x_1[i]\ (w_1[i-1], w_1[i]) + x_2[i]\ (w_1[i-1]) + z_3[i]$。从最后一个分组开始译码，然后逆向译码直到第一个分组。假设已经译码 $w_1[i]$，如果 $R_1 \leqslant \log\left(1+\frac{P_1+P_2}{N_3}\right)$，通过复用编码，那么就可以译码 $w_1[i-1]$。这种方法称为后向译码。

干扰信道在协作通信和认知无线电中起着重要的作用，其结构如图4-8所示。虽然干扰信道已经是存在了30年的开放性问题，最著名的容量域是Han&Kobayashi在1981年提出的。

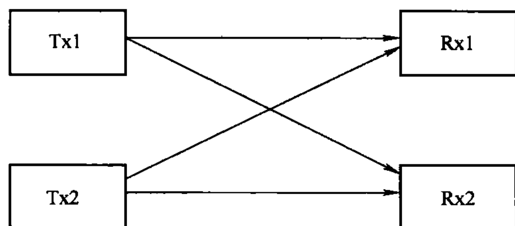


图4-8　干扰信道

4.1.2　协作分集的多址信道

本节我们将分析具有协作分集的两用户多址接入信道模型（两个协作的源节点）。两个源节点1、2产生独立的消息，这些消息编码成为码组，并且接收的消息信号和另一个消息相关。译码器观察信道输出，并且分别估计两个源节点的源消息（见图4-9）。

定义4.1：P_i 是发射机 i 的功率（$i=1, 2$）；N_i 是接收机 i 的AWGN噪声功率（$i=1, 2, 3$）；$a_{j,i}$ 是节点 i 和 j 之间的信道增益；$s_{j,i} = \|a_{j,i}\|^2\frac{P_i}{N_j}$；$\rho$ 是 X_1 和 X_2 的相关系数。

为了得到容量域的外限，我们再次使用最大流最小割理论。考虑割 S_1，可以得到

$$R_1 \leqslant \mathrm{I}(\mathrm{X}_1;\mathrm{Y}_0,\mathrm{Y}_2|\mathrm{X}_2) \leqslant \log(1+(1-\|\rho\|^2)(s_{0,1}+s_{2,1})) \tag{4-9}$$

考虑割 S_2，可以得到

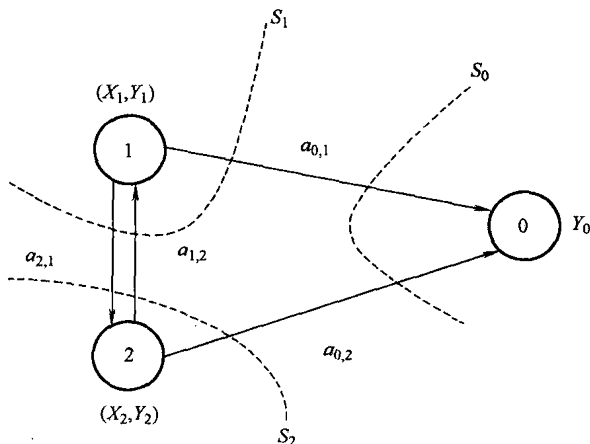


图 4-9 具有协作的多址接入模型

$$R_2 \leqslant \mathrm{I}(X_2;Y_0,Y_1|X_1) \leqslant \log(1+(1-\|\rho\|^2)(s_{0,2}+s_{1,2})) \tag{4-10}$$

考虑割 S_3，可以得到

$$\begin{aligned} R_2 \leqslant \mathrm{I}(X_1;X_2,Y_0) &\leqslant \log(1+s_{0,1}+s_{0,2}+2\cos(\angle\rho + \\ &\quad \angle a_{0,1}-\angle a_{0,2})\|\rho\|\sqrt{s_{0,1}s_{0,2}}) \\ &\leqslant \log(1+s_{0,1}+s_{0,2}+2\|\rho\|\sqrt{s_{0,1}s_{0,2}}) \end{aligned} \tag{4-11}$$

其中最后一个不等式假设发射机知道信道状态信息。

为了得到可到达容量限（容量域的内限），我们考虑以下两个方案：非协作传输和译码转发传输。

非协作传输：如果发射机不利用它们对其他发射机的观察结果，系统模型可以简化成经典的多址接入信道，因此可达到速率域为

$$\begin{aligned} R_1 &\leqslant \log(1+s_{0,1}) \\ R_2 &\leqslant \log(1+s_{0,2}) \\ R_1+R_2 &\leqslant \log(1+s_{0,1}+s_{0,2}) \end{aligned} \tag{4-12}$$

译码转发传输：编码方案的思想类似于中继信道中的方案 2（也就是译码转发 + 后向译码）。我们只给出结论：

$$\begin{aligned} R_1 &\leqslant \log(1+\alpha_1 s_{2,1}) \\ R_2 &\leqslant \log(1+\alpha_2 s_{1,2}) \\ R_1+R_2 &\leqslant \log(1+s_{0,1}+s_{0,2}+2\sqrt{(1-\alpha_1)(1-\alpha_2)s_{0,1}s_{0,2}}) \end{aligned} \tag{4-13}$$

对某些 $0\leqslant\alpha_i\leqslant1$，$i=1$，2。

非协作传输和译码转发传输的外限如图 4-10 所示，其中 $s_{0,1}=s_{0,2}=3$，$s_{2,1}=s_{1,2}=15$。

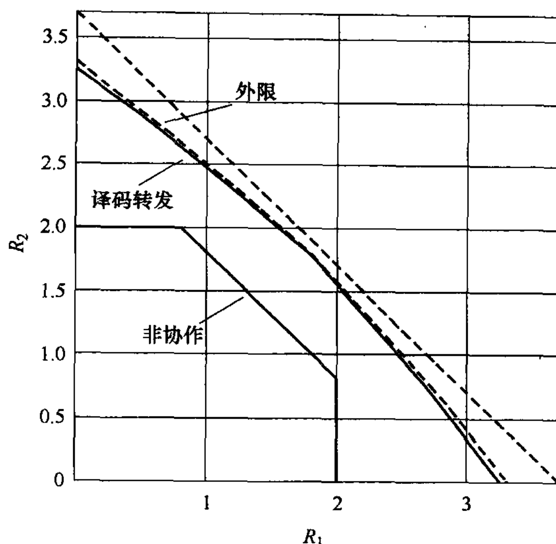


图 4-10　具有协作的多址接入网络可到达容量域

4.2　协作通信

在许多文献中已经探讨过最基本的三终端中继网络协作方案。Sendonaris 等在参考文献［8］中提出了 CDMA 系统中一个简单的协作协议。Laneman 等在参考文献［9］中提出了时分信道的一种协作分集协议。时分方式可用于很多不同的协作分集协议，例如固定中继、选择中继和增量中继。Laneman 等提出将一般的协作方案用于多终端中继网络，例如空时码协作分集协议提供多个中继的信号设计。Wittneben 等在参考文献［16，17］中提出多个中继网络中的放大转发协作分集协议。Bletsas 等在参考文献［19］中提出伺机中继，从而把多终端中继网络通过选择最佳中继终端简化成三终端中继网络。上述概念对在共存多无线电系统中推广认知无线电网络非常有意义。

4.2.1　三节点协作通信

如图 4-11 所示，在三终端中继网络中有三种基本的协作分集通信方案：译码转发、放大转发和编码协作。基站（BS）可以从两个手持终端（节点或移动台）接收信号。使用图 4-1 中的模型，一个终端是源节点，另一个终端是中继节点。基站接收到两个终端的信号来获得协作分集。在译码转发方案中，中继终端译码（或估计）从源节点发送来的信号，并且把译码后的信号转发到目的终端。如果中继终端没有正确的译码源信号，可能会造成错误的传播。在放大转发方案

中，中继终端直接把从源终端接收到的信号通过增益参数进行放大，并转发到目的节点，而不进行译码。增益参数，也称为放大参数，是用来增大信号功率。编码协作方案联合考虑了中继和源节点的信道编码和协作协议。这种方法不仅提供了协作分集，还提供了编码增益。

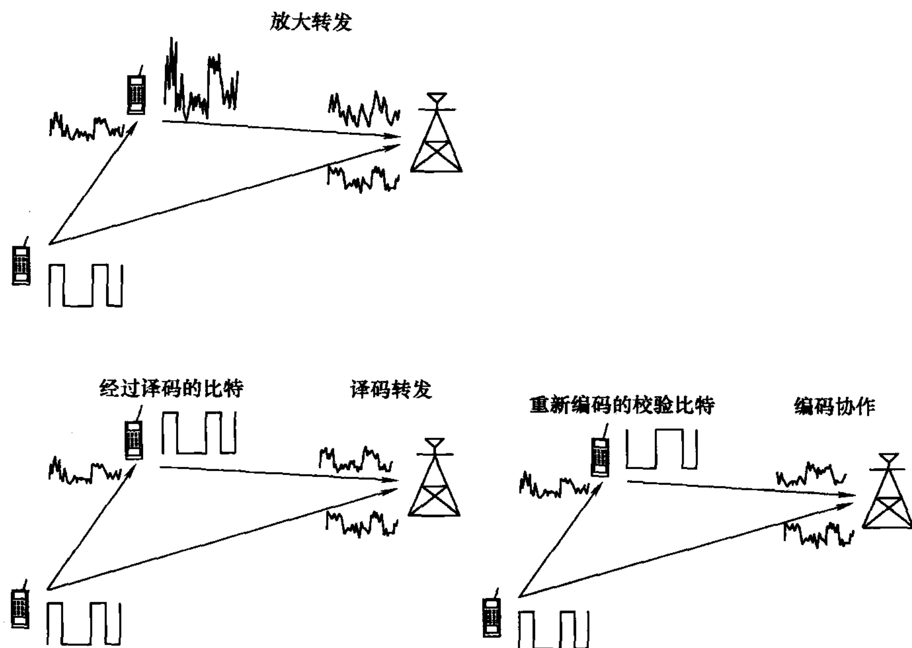


图4-11 三种基本协作分集协议[22]

4.2.1.1 CDMA协作通信网络

在图4-11所示的模型中，我们考虑两个移动台与一个终端进行通信。每个移动台发送独立的信息到目的节点，用 $W_i(i=1,2)$ 分别表示移动台1和移动台2的消息。此外，每个移动台还辅助其他移动台中继传输信号，用 $Y_i(i=1,2)$ 分别表示移动台2和移动台1到目的节点的消息。很明显，接收到的信号 Y_0 可以实现协作分集，因为每个移动台传输的信号 $X_i(i=1,2)$ 包括了 W_1 和 W_2 的信息。具体来说，在一个符号周期内基带信号模型为

$$\begin{aligned} Y_0 &= K_{10}X_1 + K_{20}X_2 + Z_0 \\ Y_1 &= K_{21}X_2 + Z_1 \\ Y_2 &= K_{12}X_1 + Z_2 \end{aligned} \tag{4-14}$$

式中，$\{K_{ij}\}$ 表示在一个符号周期内保持恒定的信道衰减系数，$Z_i(i=0,1,2)$ 分别表示目的节点、移动台1和移动台2的AWGN噪声。接收的信号 Y_0 可以分离出信号 W_1 和 W_2 来恢复信号。因此，使用两个正交信道可以很好的分离信

号：CDMA（见图 4-12）和时分方案。

例如，我们考虑使用非协作和协作通信的 CDMA 系统。假设每个用户有独立的扩频码，并且这些扩频码具有理想的正交性能。传输符号周期为 $L=3$，在一个符号周期内，每个用户传输一个信息比特。此外还假设信道相关时间比 L 个符号周期要长，因此信道衰减系数在一个符号周期内保持恒定。在非协作方案中，移动台在 3 个符号周期内传输的信号为

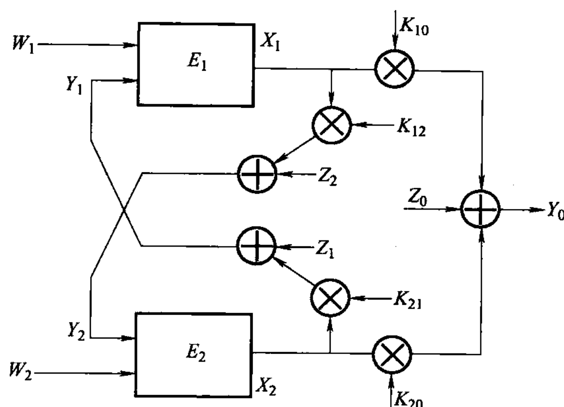


图 4-12　CDMA 信道模型

$$
\begin{aligned}
X_1(t) &= a_1 b_1^{(1)} c_1(t), a_1 b_2^{(2)} c_1(t), a_1 b_3^{(3)} c_1(t) \\
X_2(t) &= a_2 b_2^{(1)} c_2(t), a_2 b_2^{(2)} c_2(t), a_2 b_2^{(3)} c_2(t)
\end{aligned}
\tag{4-15}
$$

式中，$b_j^{(i)}$ 是移动台 j 的第 i 个比特信息，$c_j(t)$ 是移动台 j 的扩频码，a_j 是移动台 j 与功率分配相关的系数。

3 个信息比特使用不同的扩频码传输而不协作，显然，这样构造的信号并不能提供任何分集增益。接下来，我们考虑协作策略，在 3 个符号周期内移动台传输的信号是

$$
\begin{aligned}
X_1(t) &= a_{11} b_1^{(1)} c_1(t), a_{12} b_1^{(2)} c_1(t), a_{13} b_1^{(2)} c_1(t) + a_{14} \hat{b}_2^{(2)} c_2(t) \\
X_2(t) &= a_{21} b_2^{(1)} c_2(t), a_{22} b_2^{(2)} c_2(t), a_{23} \hat{b}_1^{(2)} c_1(t) + a_{24} b_2^{(2)} c_2(t)
\end{aligned}
\tag{4-16}
$$

式中，$\hat{b}_j^{(i)}$ 是辅助节点估计移动台 j 的第 i 个比特，$\{a_{ji}\}$ 是移动台 j 与功率分配相关的系数。在第一个符号周期，第一个信号比特只发送给目的节点。在第二个符号周期，信号比特发送给目的节点和辅助节点。因此，在第三个符号周期，每个移动台可以发送自己的信息比特和辅助节点的信息比特以协作传输。很明显，在协作策略中，我们可以实现协作分集，并且通过每个移动台的扩频码分离所需的信号。应该注意，在非协作方案中，每个移动台传输 3 个信息比特，而在协作方案中，每个移动台只传输 2 个信息比特，这可以证明协作方案的吞吐量比非协作方案的吞吐量要高。因此，协作方案不仅获得了协作分集，还提高了网络吞吐量。

4.2.1.2　时分信道的协作通信

为了分离源信号和中继信号，我们还需要其他正交信道分配的方案。因此时

分复用信道也可以用在协作分集中。

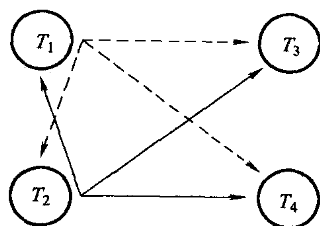


图4-13　时分协作的终端模型

时分协作的系统模型如图4-13所示。源节点、中继节点和目的节点分别用 T_s、T_r和 T_d表示，其中 s, $r\in\{1, 2\}$, $d\in\{3, 4\}$。我们考虑在频率非选择性衰落信道的窄带传输场景，因此信道衰落系数在 N 个时隙内保持恒定。图4-14给出的时分信道分配分别表示非协作和协作情况。在非协作情况（见图4-14b）的第 n 个时隙，基带信号表示为

$$y_d[n]=a_{s,d}[n]x_s[n]+z_d[n], i=1,2 \tag{4-17}$$

式中，$x_s[n]$和$y_d[n]$分别是源节点传输信号和目的节点接收信号。$\{a_{i,j}\}$是节点 i 和节点 j 的信道衰减系数，$z_i[n]$是节点 i 处的AWGN噪声。在 $n=1$, 2…, $N/2$ 时，源节点发送不带协作的源信号；另一个源节点在 $n=N/2+1$, …, N 时隙发送源信号。因此，这种非协作情况没有协作分集。在图4-14c中的协作情况时，我们把原来时长为 $N/2$ 的块分为两个长度为 $N/4$的块。在第一个时间块，源节点把源信息传输给目的节点和中继节点，中继节点监听源节点发来的源信号。在第二个时间块，中继节点把中继信号发送到目的节点。我们可以用数学建模信号，

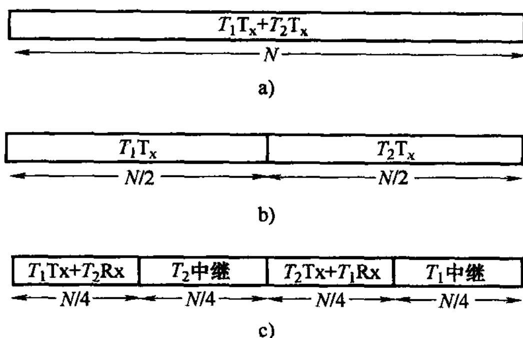


图4-14　时分信道分配
a）带干扰的直接传输　b）正交直接传输
c）正交协作传输[7]

对于 $n=1,\cdots,N/4$,

$$\begin{aligned} y_r[n]&=a_{s,r}[n]x_s[n]+z_r[n] \\ y_d[n]&=a_{s,d}[n]x_s[n]+z_d[n] \end{aligned} \tag{4-18}$$

式中，$y_r[n]$是中继节点接收到的信号。并且对于 $n=N/4+1,\cdots,N/2$,

$$y_d[n]=a_{r,d}[n]x_r[n]+z_d[n] \tag{4-19}$$

式中，$x_r[n]$是中继节点辅助源节点传输的信号。在其他符号周期内，中继节点以类似方式工作。很明显，如果我们合理地使用中继信号 $x_r[n]$，就可以达到协作分集。接下来，我们讨论能实现协作分集的这些中继方法。

4.2.1.3　固定中继

对于固定中继方法，一般有两种中继机制：放大转发和译码转发。

在放大转发方法中，中继节点的中继信号是

$$x_r[n]=\beta y_r\left[n-\frac{N}{4}\right] \tag{4-20}$$

式中，β 是放大参数。把式（4-20）代入（4-19），我们可以得到

$$\begin{aligned} y_d[n] &= a_{r,d}x_r[n]+z_d[n] \\ &= a_{r,d}\beta y_r\left[n-\frac{N}{4}\right]+z_d[n] \\ &= a_{r,d}\beta\left\{a_{s,r}x_s\left[n-\frac{N}{4}\right]+z_d\left[n-\frac{N}{4}\right]\right\}+z_d[n],n=\frac{N}{4}+1,\cdots,\frac{N}{2} \end{aligned} \tag{4-21}$$

因此，根据式（4-18），目的节点接收到的信号可以表示为

$$y_d[n]=\begin{cases} a_{s,d}[n]x_s[n]+z_d[n],n=1,\cdots,\frac{N}{4} \\ a_{r,d}\beta\left\{a_{s,r}x_s\left[n-\frac{N}{4}\right]+z_d\left[n-\frac{N}{4}\right]\right\}+z_d[n],n=\frac{N}{4}+1,\cdots,\frac{N}{2} \end{cases} \tag{4-22}$$

根据式（4-22），我们可以看出该方案达到了协作分集。注意，噪声项也被放大参数进行了增强。

在译码转发方法中，中继节点的中继信号是

$$x_r[n]=\hat{x}_s\left[n-\frac{N}{4}\right],n=\frac{N}{4}+1,\cdots,\frac{N}{2} \tag{4-23}$$

把式（4-23）代入式（4-19），可以得到

$$\begin{aligned} y_d[n] &= a_{r,d}x_r[n]+z_d[n] \\ &= a_{r,d}\hat{x}_s\left[n-\frac{N}{4}\right]+z_d[n] \end{aligned} \tag{4-24}$$

因此，根据式（4-18），目的节点接收到的信号可以表示为

$$y_d[n]=\begin{cases} a_{s,d}[n]x_s[n]+z_d[n],n=1,\cdots,\frac{N}{4} \\ a_{r,d}\hat{x}_s\left[n-\frac{N}{4}\right]+z_d[n],n=\frac{N}{4}+1,\cdots,\frac{N}{2} \end{cases} \tag{4-25}$$

因此，通过中继节点估计源信号可以实现协作分集。注意，如果估计错误较大，协作分集的性能将会变差。

4.2.1.4　固定中继的中断行为

在移动通信中除了链路信噪比，还需要判断衰落信道的中断性能。因此我们考虑慢衰落下的中断概率和中断容量，由于在慢衰落中信道延迟限制与信道相关时间有关。在 AWGN 信道和慢衰信道中概念上不同，因此我们在慢衰信道中使用中断容量而不是香农容量。在 AWGN 信道中，用户可以正速率发送数据，并

且数据传输差错率任意小。而对慢衰信道，只要信道处于深衰落的概率不为0，那么用户不能总是以正速率发送数据。因此，严格来说，慢衰信道的香农容量为0，这也是我们为什么用中断容量取代香农容量分析的原因。中断概率 $P_{out}(R):=P\{\log(1+\|h\|^2 SNR)<R\}$ 是信道不能支持传输速率 R 的概率。中断容量是中断概率 $P_{out}(R)$ 低于 ε(预先定的值）的最大传输速率 R。此外，我们假设只有在接收机能获得信道状态信息，并且半双工传输，也就是节点不能同时发送和接收数据。下面的讨论是在近似高信噪比的前提下基于正交直接传输和中继传输的中断概率。

直接传输：根据式（4-14），最大互信息是 $I_D=\log\ (1+SNR\|a_{s,d}\|^2)$。频谱效率 R 的中断事件是 $I_D<R$。对于瑞利衰落，也就是 $\|a_{s,d}\|^2$ 服从参数为 $\sigma_{s,d}^{-2}$ 的指数分布，中断概率是

$$P_D^{out}(SNR,R):=P_r[I_D<R]=P_r\left[\|a_{s,d}\|^2<\frac{2^R-1}{SNR}\right]\sim\frac{1}{\sigma_{s,d}^2}\cdot\frac{2^R-1}{SNR}$$

注意，分集增益（中断概率的信噪比部分）是1。

放大转发：中继节点把接收信号进行放大：

$$x_r[n]=\beta y_r\left[n-\frac{N}{2}\right]\quad n=N/2+1,\cdots,N$$

$$\beta\leqslant\sqrt{\frac{P}{\|a_{s,r}\|^2+N_0}}$$

把式(4-15)代入式(4-17)，可以得到

$$\begin{bmatrix}y_d[n]\\y_d[n+N/2]\end{bmatrix}=\begin{bmatrix}a_{s,d}\\a_{r,d}\beta\, a_{s,r}\end{bmatrix}x_s[n]+\begin{bmatrix}0&1&0\\a_{r,d}\beta&0&1\end{bmatrix}\begin{bmatrix}z_r[n]\\z_d[n]\\z_d[n+N/2]\end{bmatrix}\quad n=1,\cdots,N/2$$

组成了矢量高斯信道，可以得到互信息为

$$I_{AF}=\log(1+SNR\|a_{s,d}\|^2+f(SNR\|a_{s,r}\|^2,SNR\|a_{r,d}\|^2))$$

式中，$f(x,\ y):\ =\frac{xy}{x+y+1}$，因此中断概率为

$$P_{AF}^{out}(SNR,R):=P_r[I_{AF}<2R]\sim\left(\frac{1}{2\sigma_{s,d}^2}\ \frac{\sigma_{s,r}^2+\sigma_{r,d}^2}{\sigma_{s,r}^2\sigma_{r,d}^2})(\frac{2^{2R}-1}{SNR}\right)^2$$

（为了与直接传输的频谱效率形式保持一致，R 增大了2倍。）注意，放大转发的分集增益为2。

译码转发：中继节点完全译码 $y_r[n]n=1,\ \cdots,N/2$，重新编码（根据重复编码方案）并且重传源消息：$x_r[n]=\hat{x}_s\left[n-\frac{N}{2}\right]$，$n=N/2+1,\ \cdots,N$。互信息为

$$I_{DF}=\min\{\log(1+SNR\|a_{s,r}\|^2),\log(1+SNR\|a_{s,d}\|^2+SNR\|a_{r,d}\|^2)\}$$

式中，第一项表示中继节点准确译码源消息的最大速率，而第二项表示目的节点可以准确译码重复传输的源消息的最大速率。中断概率为

$$P_{DF}^{out}(SNR,R):=P_r[I_{DF}<2R]\sim\frac{1}{\sigma_{s,d}^2}\frac{2^{2R}-1}{SNR}$$

可以看出译码转发的分集增益为 1。这就表明固定译码转发在较大信噪比时不能提供分集增益，因为中继节点需要译码源信息，从而限制了系统性能。

4.2.1.5　**选择中继和增量中继**

在上面提到的五种中继方法中，中继节点把源节点到目的节点的源信息进行中继。然而，当中继节点和目的节点的信道状况比源节点和目的节点的信道要差时，简单地把源节点的源信道重复传输到目的节点可能比在中继时段从中继节点把信号中继到目的节点要好。因此，选择中继方法需要根据信道状况 $|a_{s,r}|^2$ 判断是否传输中继信道到目的节点。具体来说，如果

$$|a_{s,r}|^2<D \tag{4-26}$$

那么中继信道不把信号中继到目的节点，其中 D 是固定门限。由于目的节点总能在第二个时间段接收到从源节点和中继节点发送来的重复源信号，所以能实现协作分集。

根据链路状况的反馈信息，增量中继比其他中继方法能更加有效地使用中继节点。反馈信息包括传输是否成功的信息，并且源节点和中继节点都能获得反馈信息。因此源节点和中继节点能判断是从源节点还是从中继节点接收信号。很明显，这种方法比选择中继方法更有效，因为只有一个选择（接收或中继），因此减少了接收信号所需的时间。

4.2.2　多节点中继网络

在很多通信场景中，例如蜂窝网络，可能有超过三个节点存在。因此，我们可能需要把一个中继节点的场景扩展到多个中继节点，这样可以提供更大的协作分集。为了更好地理解多中继节点的概念，在下面的章节中我们将讨论三种代表性的多中继节点方案。

4.2.2.1　**空时码协作**

如图 4-15 所示，该多中继节点模型由两阶段组成。在阶段 I，源节点把源信号传输到目的节点和所有的中继节点。因此，所有中继节点在阶段 I 监听源信号。在阶段 II，中继节点使用空时码，从而在目的节点可以把中继信号进行

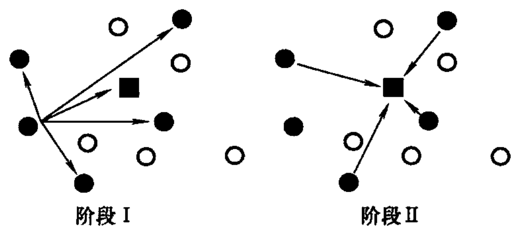


图 4-15　多中继节点模型

合并。

此外，为了把源信号和中继信号分离开，我们把信道划分为如图4-16所示的多个正交子信道。协作节点集用 $M=\{1, 2, \cdots, m\}$ 表示。对源节点 s，通过空时码辅助源节点 s 的中继节点集用 $D(s)$ 表示，与源节点 s 对应的目的节点用 $d(s)$ 表示。

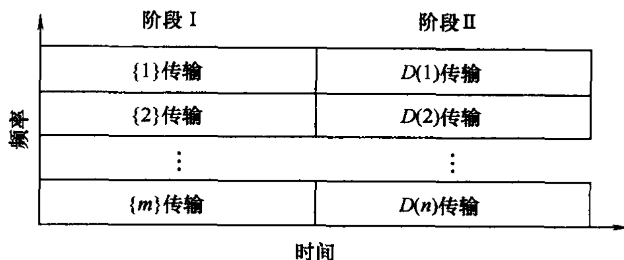


图4-16 信道分配

我们考虑阶段Ⅰ和阶段Ⅱ时间周期为 N 的基带模型。每个阶段时间长度为 N 的一半。在阶段Ⅰ中 $n=1, \cdots, N/2$，每个中继节点 $r\in M-\{s\}$ 接收到信号

$$y_r[n]=a_{s,r}[n]x_s[n]+z_r[n] \tag{4-27}$$

式中，$x_s[n]$ 和 $y_r[n]$ 分别表示源节点传输的信号和中继接收的信号；$\{a_{i,j}\}$ 是节点 i 和节点 j 的信道衰落系数；$z_i[n]$ 是节点 i 的 AWGN 噪声。目的节点 $d(s)$ 也接收到信号

$$y_{d(s)}[n]=a_{s,d(s)}[n]x_s[n]+z_{d(s)}[n] \tag{4-28}$$

在阶段Ⅱ中 $n=N/2+1, \cdots, N$，目的节点 $d(s)$ 接收到信号

$$y_{d(s)}[n]=\sum_{r\in D(s)}a_{r,d(s)}[n]x_r[n]+z_{d(s)}[n] \tag{4-29}$$

式中，$x_r[n]$ 表示中继节点 r 传输的中继信号。这些中继信号通过空时码来实现协作分集。

4.2.2.2 放大转发协作

图4-17所示的多中继节点模型可以用于一般的互连场景，例如 Ad hoc 网。源节点和目的节点分别用 S_q 和 D_q 表示，假设源节点和目的节点的数量都等于 N_a。中继节点用 R_j 表示，中继节点数量为 N_r。放大转发协作协议也由两个阶段组成。在阶段Ⅰ，源节点把源信号传输到中继节点。在阶段Ⅱ，中继节点放大并转发接收到的信号到目的节点。

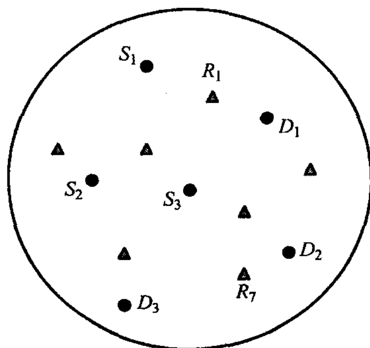


图4-17 多中继协作模型

放大转发多中继网络的系统模型如图4-18

所示。这些源节点传输的信号用信号矢量 $\vec{s}=[s_1, s_2, \cdots, s_{N_a}]^T$表示，其中元素$s_i$表示第 i 个源节点的源信号。目的节点对应的接收信号用$\vec{y}=[y_1, y_2, \cdots, y_{N_a}]^T$表示，其中元素 y_i是第 i 个目的节点的接收信号。$H_{sr}[j, q]$ 是源节点 q 到中继节点 j 的信道衰落系数，因此 $\boldsymbol{H}_{sr}$是源节点和中继节点的信道矩阵。$\vec{m}$ 是中继节点的 AWGN 噪声。$\boldsymbol{D}_r$是对角增益矩阵，其中元素 d_j表示放大转发方案的放大参数。$\boldsymbol{H}_{rD}$是中继节点和目的节点的信道矩阵。图 4-18 的信号模型等效于图 4-19，其中 $\boldsymbol{H}_{SD}$是源节点和目的节点的等效信道矩阵。显然，如果 $\boldsymbol{H}_{SD}$是对角阵，这种等效模型中不存在干扰。由于这种等效信道矩阵与对角增益矩阵 D_r有关，我们可以通过设计放大参数使得 $\boldsymbol{H}_{SD}$ 为对角阵。因此，可以有效的实现协作分集。

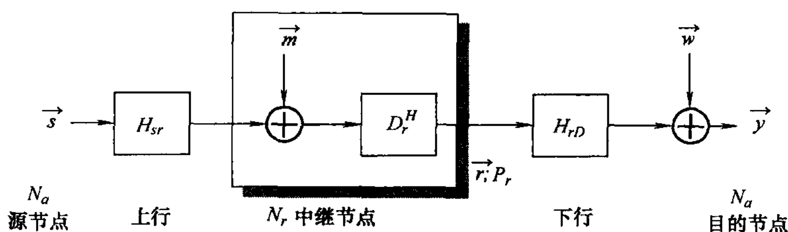


图 4-18　多中继互连的系统模型

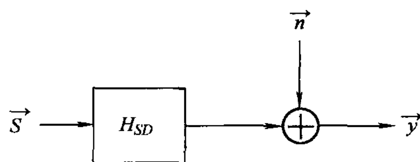


图 4-19　等效系统模型

4.2.2.3　伺机中继

在多中继环境下，所有的中继节点可以辅助源节点中继源信号。然而，我们可以选择最佳的中继节点来中继信号。伺机中继的基本思路如图 4-20 所示，我们仍然考虑两阶段方案。假设只有一个源节点和一个目的节点，以及一系列可能的中继节点。在阶段 I，源节点把源信号传输到目的节点和所有中继节点。在阶段 II，所有的中继节点把从源节点接收到的信号用之前章节介绍的协作方法，例如空时码的方式，中继到目的节点。为了减少这些中继节点的负荷，伺机中继方法选择出最优的中继路径，通过信道测量只选择一个中继节点，从而保证了链路性能。

伺机中继的概念可以推广到 Ad hoc 网络中的认知无线电和现存的多无线电系统。

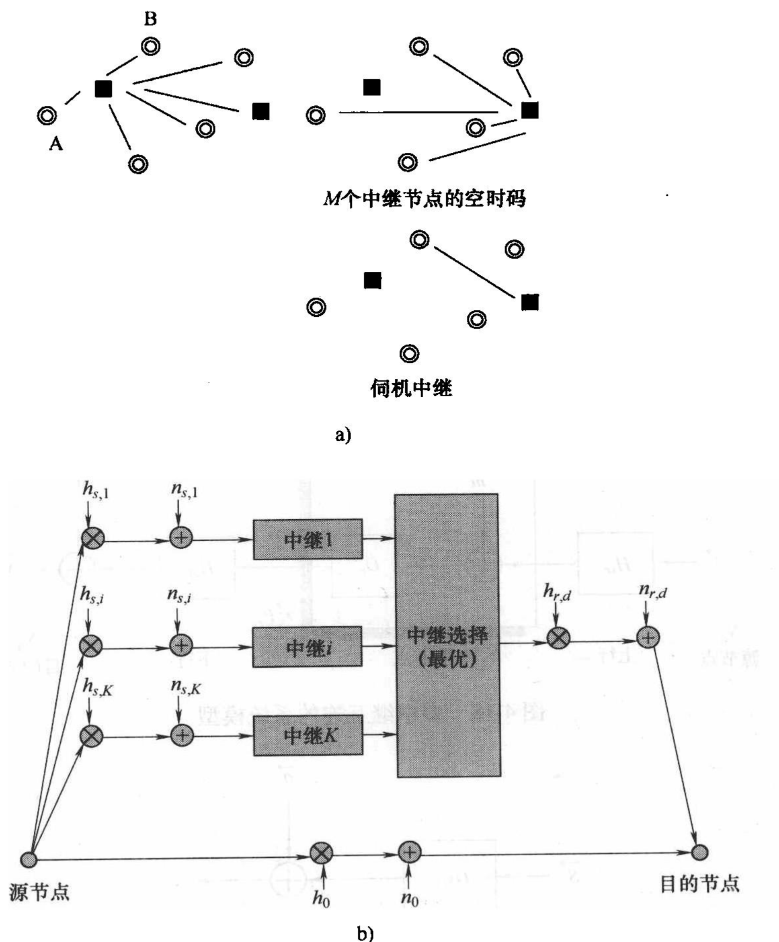


图 4-20 伺机中继

a) 终端链路模型 b) 中继选择

具体来说，用 $a_{s,i}$和 $a_{i,d}$分别表示源节点和中继节点 i、中继节点 i 和目的节点之间的信号衰减系数，其中信道假设是频率非选择性衰落信道。在伺机中继中，中继节点 i 需要知道 $a_{s,i}$和 $a_{i,d}$的信息。有两种常见的选择最优中继节点的策略：

中继选择策略 A：

$$h_i = \min\{\,|a_{s,i}|^2,\,|a_{i,d}|^2\} \tag{4-30}$$

中继选择策略 B：

$$h_i = \frac{2}{\dfrac{1}{|a_{s,i}|^2} + \dfrac{1}{|a_{i,d}|^2}} \tag{4-31}$$

式中，h_i是选择中继节点 i 的信道测量（或信道增益）。很明显，式（4-30）和式（4-31）为每个中继节点 i 提供了源-中继-目的路径的信道测量。因此，通过上述中继选择策略，减低了多中继节点的负荷，并且成功地获得协作分集。

4.3　协作无线网络

到这一节为止，协作通信一般是看成是物理层的一种空间分集，因为经过多个独立衰落信道传输的信号可以为目的节点提供更多信息，即使每个节点只有单根天线。本节我们主要讨论中继选择的一些网络场景。选择中继的想法最初是在参考文献［7］中提出的，在参考文献［4］中用于分布式空时码（DSTC）协议。但是随着中继数量的增加，之前的方法变得不可行，因此在参考文献［13］中提出了中继选择的方法。虽然这两种方法都能实现全分集，我们的目标是用合理的代价提供更大的分级增益。也就是，我们的方法等效于在目的节点部署多根天线。但是，如图4-21所示，中继选择概念可以在 Ad hoc 网络中起到更大的作用。

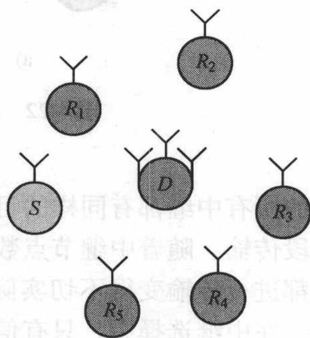
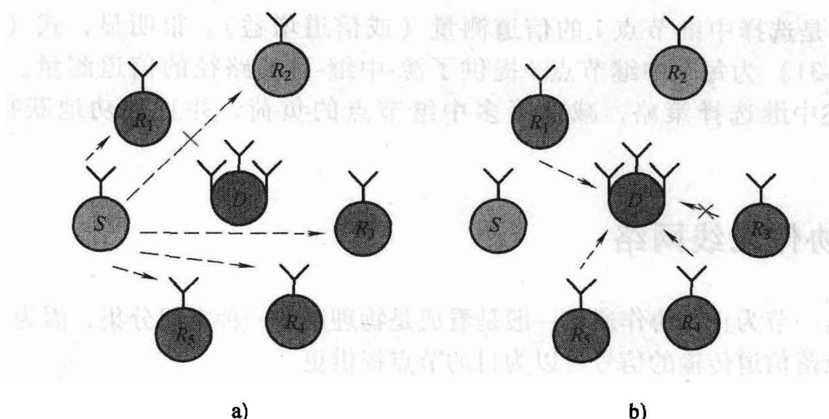


图4-21　$m=5$ 的协作中继 Ad hoc 网络拓扑

4.3.1　无线网络中协作的好处

图4-21 中的 Ad hoc 网由单个源节点、单个目的节点和 $m=5$ 个中继节点组成。源节点可以和目的节点直接通信。一共有 m 个固定中继节点，分别用 R_1，R_2，…，R_m表示。中继节点使用译码转发（DF）方法中继。假设两个节点 i 和 j 之间的信道是方差为 $\sigma_{i,j}^2$的慢衰瑞利信道，并且假设所有无线电设备都工作在半双工模式，承载信息的符号传输分为两个阶段：广播阶段和协作阶段。在广播阶段，源节点发送信号到中继节点和目的节点。在广播之后，每个中继节点译码接收到的信号。在成功译码之后，中继节点加入译码集 $D(S)$，其中 S 表示源节点。两个阶段的操作如图4-22所示，其中假设一共有5个中继节点。在阶段1，除了 R_2之外的所有中继节点译码从源节点接收到的源信号。因此 R_2不能转发信道到目的节点。在下一个阶段，成功译码源信号的中继节点可以转发译码后的信号。然而，R_3和目的节点之间的信道质量较差，因此不能成功传输。

在确定译码集 $D(S)$ 之后，就可以区分选择用于中继的中继节点。当 $D(S)$



a) b)

图 4-22 具有中继选择的协作 Ad hoc 网络两阶段

a）广播阶段 b）协作阶段

中的所有中继都有同样的目的节点时，这些中继节点在不同的频带或者在不同的时段传输。随着中继节点数量的增加，对系统的需求也随之提高，因此让所有中继都进行传输变得不切实际。中继选择是一种能够降低规模大的系统需求的方法。在中继选择中，只有信道状况最好的一个中继节点会被选择进行中继，并且与原始的方案一样可以实现分集阶数 m。

为了进一步提高协作网络的分集阶数，我们让目的节点配置上多天线。显然这种方案可以只增加一些成本而取得更大的分集增益，成本包括物理天线和反馈负荷。随着 MIMO 技术逐渐成熟，配置多根天线的设备将不再成本很高，而是价格比较合理。目的节点选择中继节点所需的反馈负荷是 $\log_2(\min(J, M))$，反馈负荷并不随着中继节点数量增加而无限制的增长。

假设在一个时隙内传输所需的端到端数据速率为 R，考虑到中继的特性和半双工的工作方式，中继路径上一共有两跳。在每一跳上，至少需要容量为 $2R$。对第一跳，如果容量小于 $2R$，中继不能成功译码源信号，因此不能中继信号。如果中继到目的节点的容量小于 $2R$，那么目的节点不能成功译码。也就是对中继节点 j

$$\frac{1}{2}\log_2\left(1+|h_{Rj,R}|^2\frac{P}{N_0}\right)=\frac{1}{2}\log_2(1+|h_{Rj,R}|^2SNR)\geqslant R\ \frac{1}{2}\log_2(1+|h_{Rj,D}|^2SNR)\geqslant R$$

式中，$SNR=P/N_0$，P 是每个节点的最大传输功率，N_0是噪声功率谱密度。定义中断概率 P_{out}为源节点和目的节点之间的互信息小于所需的传输速率 R 的概率。在我们的场景中，中断事件是在 $\mathrm{I}_{sel}<R$ 时发生，其中 I_{sel}是使用中继选择的源节点和目的节点之间的互信息，可以表示为

$$\mathrm{I}_{sel}=\frac{1}{2}\log_2\left(1+|h_{S,D}|^2SNR+\frac{HKH^*}{N_0}\right) \tag{4-32}$$

式中，$K=\text{diag}(P_1,\ \cdots,\ P_m)$ 是选择的每个中继传输功率，为了简化分析，假设都等于 P，$H=[h_1^{sel},\ h_2^{sel},\ \cdots,\ h_m^{sel}]$ 表示选择的 m 个矢量的复信道矩阵，其中每个矢量是从中继节点到目的节点的接收天线阵列的信道冲激响应。注意 h_i^{sel} 是在 $h_{R_1,D}$，$h_{R_2,D}$，$\cdots$，$h_{R_j,D}$ 中第 i 个最好的信道，其中 $h_{R_j,D}=[h_{R_1,1},\ h_{R_2,2},\ \cdots,\ h_{R_j,M}]^T$ 表示从中继节点 j 到目的节点的信道冲激矢量，可以由下式得到：

$$h_i^{sel}=h_j/\max_{j\in\{1,\cdots k\}}\{h_1^{sel},\cdots,h_{i-1}^{sel}\}h_jh_j^*$$

因此

$$Pr[\text{I}_{sel}<R]=\sum_{D(S)}Pr[\text{I}_{sel}<R\mid D(S)]Pr[D(S)]$$

属于译码集的概率 $Pr[R_j\in D\ (S)]$：根据参考文献 [7]，如果一个中继节点能满足下列条件，那么该中继可以属于译码集。

$$Pr[R_j\in D(S)]=Pr\left[|h_{S,R_j}|^2>\frac{2^{2R}-1}{SNR}\right]=\frac{1}{2\sigma^2_{S,R_j}}\exp\left(-\frac{1}{2\sigma^2_{S,R_j}}\frac{2^{2R}-1}{SNR}\right)$$

假设中继路径相互独立

$$\begin{aligned}Pr[D(S)]&=\prod_{R_j\in D(S)}\exp\left(-\frac{1}{2\sigma^2_{S,R_j}}\frac{2^{2R}-1}{SNR}\right)\\&\cdot\prod_{R_j\notin D(S)}\left(1-\exp\left(-\frac{1}{2\sigma^2_{S,R_j}}\frac{2^{2R}-1}{SNR}\right)\right)\approx\left(\frac{2^{2R}-1}{SNR}\right)^{K-D(S)}\\&\cdot\prod_{R_j\notin D(S)}\frac{1}{2\sigma^2_{S,R_j}}\qquad(SNR\to\infty)\end{aligned}$$

译码集的条件中继概率 $Pr[\text{I}_{sel}<R|D(S)]$：其中我们一共有 $\min(J,\ M)$ 个中继节点，为了简化分析，我们假设 $J>M$，在实际无线网络中该假设也是合理的。为了得到条件概率，我们必须找到选择中继的互信息的表达式。首先，我们定义了三个随机变量 X、Y_i、Z_{ij} 为

$$X=\sum_{i=1}^{M}Y_i=\sum_{i=1}^{M}\sum_{j=1}^{M}Z_{i,j}$$

式中，Z 是信道衰落的二次方值，Z 的分布满足

$$f_z(z_{i,j})=|h_{i,j}|^2\sim\frac{1}{2\sigma^2_{i,j}}\exp\left(-\frac{1}{2\sigma^2_{i,j}}Z_{i,j}\right)$$

对 $i\in\{1,\cdots,K\}$，Y_i 是 M 个随机变量 $Z_{i,j}$ 对 j 求和，X 是在 Y_i 中 M 个最大的随机变量的和。一旦找到 X 的分布，中继选择的互信息可以表示为

$$\text{I}_{sel}=\frac{1}{2}(1+SNR(|h_{S,D}|^2+X))\tag{4-33}$$

根据互信息，可以很容易得到中断概率 $Pr\ [\text{I}_{sel}<R]$。

除了中断事件之外，我们还关心 I_{sel} 是否提高了相应的端到端容量。类似的，

一旦得到 X 的分布，可以得到 I_{sel}。在无线（Ad hoc）网络中获得 I_{sel} 具有重要的意义，因为我们可以通过协作增大累积带宽，而不是增加分集以对抗信道衰落。

4.3.2 基于分簇 Ad hoc 网络中的协作

在很多（移动）Ad hoc 网络中（例如 MANET），节点通过簇头或接入点（AP）进行通信。此外，AP 可能连接到骨干网来组成混合基础设施架构和 Ad hoc（HIAH）网络。在 MANET 中，AP 可以收集网络节点的状态信息，例如信道衰减系数，从而根据性能准则选择出最合适的协作模式，并且通过合适的控制信道（或数据包）反馈 AP 的决定。

网络架构如图 4-23 所示，类似于两层网络，在簇与簇之间提供了更可靠和长距离的连接[21]。对于 AP，匹配链路和激活协作链路更具有挑战。典型的分簇 MANET 实现的方法可能包括下列步骤。

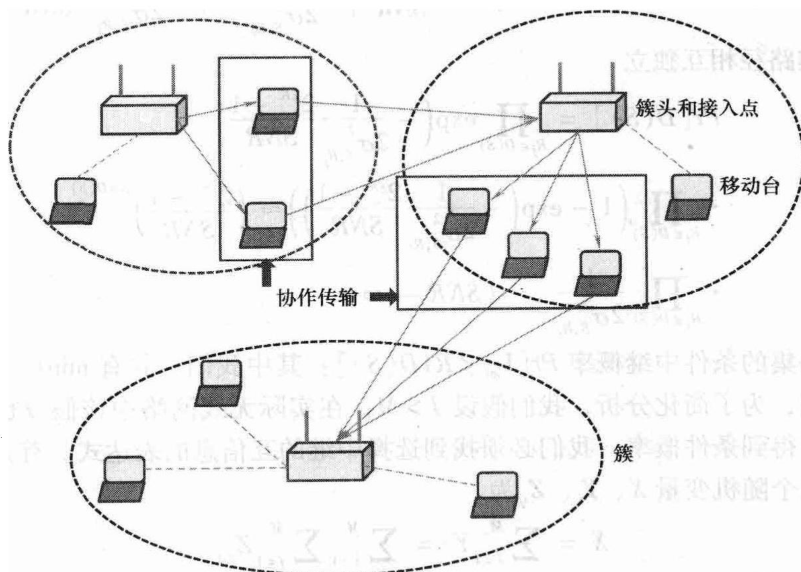


图 4-23 直接传输和协作传输的簇

1）划分基础设施网络：我们把节点（或移动终端）分为协作对，这可以看成是图论里的匹配问题。最小权重匹配算法是寻找具有最小权重的匹配，其复杂度是节点数量的三次方。贪婪匹配算法则是通过随机选择自由节点，并与其他节点中最好的节点进行匹配，该算法的复杂度是节点数量的四次方。我们还可以随机匹配节点，此时复杂度与节点数量是线性关系。当然，低复杂度的匹配算法中断概率较高。

2）协作连接：当通过无线网络传输累积信息时，应该找出利用网络容量的

连接性（也就是协作节点对）。如果我们假设已知路径损耗和衰落统计特性，如果接收到的信噪比在给定的门限之上，那么该链路是可行的。有限节点随机分布在二维区域的 Ad hoc 网络的连接性已经被深入的研究。假设路径损耗是两个节点 i 和 j 的距离 d_{ij} 的单调递增函数（也就是 $d_{ij}^{-\alpha}$），并且用 A 表示节点 i 为中心的圆形区域，其中对所有节点 $j \in A$ 都有 $SNR(d_{ij}) \propto d_{ij}^{-\alpha}$，信噪比在门限之上，从而保证连接性 $SNR(d_{ij}) > \eta$，任何两个相互距离比 A 的半径小的两个节点都能保证互连。在任何两个距离小于 A 的半径之间的点画直线得到的图显示了那些节点能直接相互通信，以及能通过多跳路径相互通信的节点集合，这个集合称为簇。

在这种设置上，有两种对网络连接不同的定义。在稀疏网络设置中，如果网络中出现一个包括无限数量节点组成的簇（称为无限簇），该网络定义成已连接的。在密集网络设置中，如果所有的节点对可以相互通信，那么该网络可以定义为正在连接。我们把后者定义称为网络完全连接。两种定义都是网络图本质的特性，并且在分簇多跳网络里节点图的边被用于通信。在大规模的稀疏网络分析中，节点一般以密度 λ 分布在无限的二维平面。通过过滤理论可以分析网络连接性。很明显，增大 λ 可以增强网络的连接性。对于密集网络，一般分析随机分布在单位区域平面的 N 个节点。Gupta 和 Kumar 研究了在单位面积分布了较多的 N 个节点场景下保证网络全连接性的充分必要条件：每个节点无线电的覆盖范围必须大于 $A^{\text{non-coop}} = N^{-1}[\log N + c(N)]$，其中 $\liminf c(N) = +\infty$。这个条件是必须的，可以证明节点的覆盖范围小于 $N^{-1}N[\log N + c(N)]$（其中 $\limsup c(N) < +\infty$）的网络几乎不可能完全连接。

3）协作分簇网络的连接性：在协作网络中，分簇可以通过增大搜寻新邻居节点的范围来增强连接性，如图 4-24 所示。事实上，如果不是所有的节点都孤立，那么在 AP 搜寻新的邻居节点时一定能找到已连接簇内的节点。在密集网络中，可以证明不满足非协作网络全连接必要条件的协作网络可以较大概率完全连接。证明过程主要是由于分簇把网路区域分为小的部分，所有这些小区域都有较大概率在区域 A 内节点组成很多簇，还有可能包括在临近区域的簇，因此整个网络可以保持全连接。协作分簇网络连接性中节点所需要覆盖的范围是

$$A^{\text{coop}} \geqslant N^{-1} 4\pi (4\log N)^{\frac{\alpha}{\alpha+2}} (\log\log N + \log 2)^{\frac{2}{\alpha+2}} \tag{4-34}$$

协作 Ad hoc 网络中保持连接性所需的功率增益为$\left(\frac{\log N}{\log\log N}\right)^{\frac{2}{\alpha+2}}$。

实际上，除了上面描述的方法之外，我们可以使用随机协作编码和随机分簇来随机化协作，从而提高网络连接性。在任何情况下，协作不限制于点对点通信，而是一种基本的网络功能。

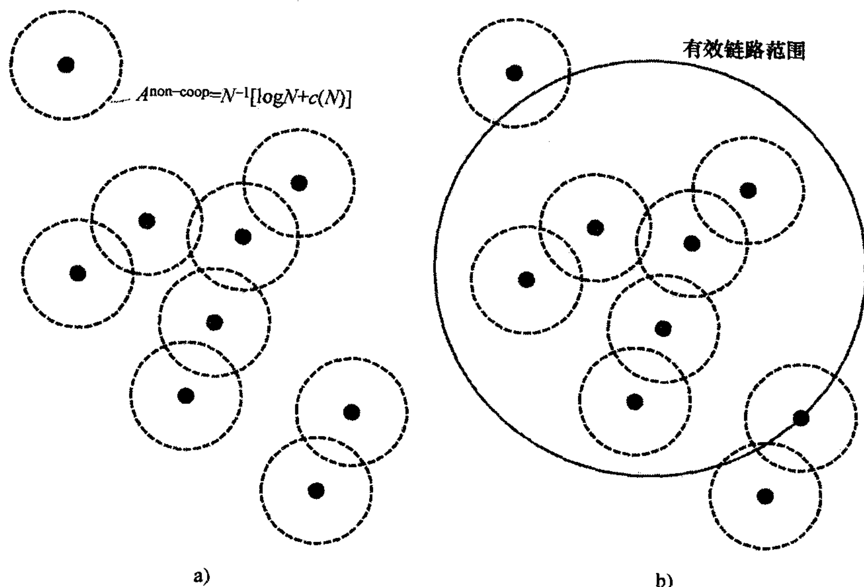


图 4-24 协作簇的连接性

参考文献

[1] T. Cover, J. Thomas, *Information Theory*, John Wiley & Sons Inc., New York, 1991.

[2] D. Tse, *Fundamentals of Wireless Communication*, Cambridge University Press, Cambridge, UK, 2005.

[3] T.M. Cover, A.A.E. Gamal, 'Capacity Theorems for the Relay Channel', *IEEE Trans. Info. Theory*, **25**(5), 1979, 572–584.

[4] J. Laneman, G. Wornell, 'Distributed Space-time-coded Protocols for Exploiting Cooperative Diversity in Wireless Networks', *IEEE Trans. Inf. Theory*, **49**(10), 2003, 2415–2425.

[5] J.N. Laneman, 'Cooperative Diversity in Wireless Networks: Algorithms and Architectures', Ph.D. dissertation, MIT, Cambridge, MA, 2002.

[6] J.N. Laneman, G.W. Wornell, 'Distributed Space-time-coded Protocols for Exploiting Cooperative Diversity in Wireless Networks', *IEEE Transactions on Information Theory*, **49**(10), 2003, 2415–2425.

[7] J.N. Laneman, D.N.C. Tse, G.W. Wornell, 'Cooperative Diversity in Wireless Networks, Efficient Protocols, and Outage Behavior', *IEEE Trans. Inf. Theory*, **50**(12), 2004, 3062–3080.

[8] A. Sendonaris, E. Erkip, B. Aazhang, 'User Cooperation Diversity Part I and Part II', *IEEE Trans. Commun.*, **51**(11), 2003, 1927–1948.

[9] J.N. Laneman, G.W. Wornell, D.N.C. Tse, 'An Efficient Protocol for Realizing Cooperative Diversity in Wireless Networks', *Proc. IEEE ISIT*, Washington, DC, June 2001, p. 294.

[10] J.N. Laneman, G.W. Wornell, 'Exploiting Distributed Spatial Diversity in Wireless Networks', in *Proc. Allerton Conf. Communications, Control, and Computing*, Monticello, IL, Oct. 2000.

[11] T.E. Hunter, A. Nosratinia, 'Cooperative Diversity through Coding', *Proc. IEEE ISIT*, Lausanne, Switzerland, *July* 2002, p. 220.

[12] A. Høst-Madsen, 'Capacity Bounds for Cooperative Diversity', ' *IEEE Trans. Inf. Theory*, **52**(4), 2006, 1522–1544.

[13] E. Beres, R. Adve, 'Selection Cooperation in Multi-Source Cooperative Networks', *IEEE Transactions on Wireless Communications*, **7**(1), 2008, 118–127.

[14] T.E. Hunter, A. Nosratinia, 'Diversity through Coded Cooperation', submitted to *IEEE Trans. Wireless Commun.*, 2004.

[15] J.N. Laneman, G.W. Wornell, 'Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks', *IEEE Trans. Inform. Theory*, **49**,October 2003, 2415–2525.

[16] A. Wittneben, B. Rankov, 'Distributed Antenna Systems and Linear Relaying for Gigabit MIMO Wireless', *IEEE Vehicular Technology Conference VTC 2004 Fall, Los Angeles, USA*.

[17] A. Wittneben, I. Hammerström, 'Multiuser Zero Forcing Relaying with Noisy Channel State Information', *IEEE WCNC*, 2005.

[18] A. Bletsas, A. Khisti, D.P. Reed, A. Lippman, 'A Simple Cooperative Diversity Method Based on Network Path Selection', *IEEE JSAC*, 2006.

[19] A. Bletsas, H. Shin, M.Z. Win, A. Lippman, 'Cooperative Diversity with Opportunistic Relaying', *IEEE WCNC*, 2006.

[20] A. Nosratinia, T.E. Hunter, A. Hedayat, 'Cooperative Communication in Wireless Networks', *IEEE Communications Magazine*, **42**(10), 2004, 74–80.

[21] A. Scaglione, D.L. Goeckel, J.N. Laneman, 'Cooperative Communications in Mobile Ad Hoc Networks: Rethinking the Link Abstraction', *IEEE Signal Processing Magazine*, September 2006, 18–29.

[22] Raymond W. Yeung, A First Course in Information Theory, Kluwer, 2006.

第5章　认知无线电通信

5.1　认知无线电和动态频谱接入

正如在第2章介绍的，随着无线通信的迅速发展，在具有使用许可和无需使用许可频带上有许多不同的通信系统用于满足不同的需求和应用，例如GSM/GPRS、IEEE802.11、蓝牙、UWB、Zigbee、3G（CDMA系列）、HSPA、3G LTE、IEEE 802.16等。另一方面，由于无线电波非视距传播的特性，一般无线设备都更倾向于使用3GHz以下的频段。因此，越来越多的设备，到2020年估计有1万亿无线设备需要分配无线频谱，这也是对未来无线通信最大的挑战。

在过去，频谱分配是对特定的服务分配特定的频带。美国联邦通信委员会（FCC）使用的频率分配图如图5-1所示[3]。该频谱分配图看上去具有很高的频谱利用率。但是，FCC频谱策略任务组在参考文献[4]中指出，适合非视距无线传播的3GHz以下频段中已经分配的频段在不同的时间和地点只有15%～85%正

图5-1　FCC频谱分配图

在被使用。3GHz 以上频段的频谱利用率更低，换言之，已经分配的频谱中很大一部分只是偶尔被使用，从而导致很大一部分的频谱利用率很低。

虽然固定频谱分配政策在过去起到了重要的效果，但是随着近年来要使用有限频谱资源的移动业务和应用的急剧增长，传统固定分配的频谱政策正限制着移动业务的发展。由于无线传播特性导致可用的频谱十分有限，对频谱使用更高效率的需求推动了高效使用现有频谱资源的通信新技术。随着在 2.4GHz 免使用许可 ISM 频带和其他频带上多种无线技术成功地在全球共存使用，动态频谱接入作为一种解决目前频谱使用效率低下问题的方法被提出。通过伺机接入需要频率使用许可的频段和与现有用户（主用户）共享使用频率，可以有效地提高现有频谱的使用效率。

动态频谱接入最关键的技术是认知无线电（CR）技术，认知无线电提供了与授权用户用伺机的方法共享无线信道的能力。认知无线电能够通过异构无线架构和动态频谱接入技术为移动用户提供高带宽。由于可用频谱范围较宽，另外不同的应用需要不同的服务质量，所以认知无线电网络中存在许多技术挑战。

为了在不干扰授权用户的前提下与授权用户共享频谱，并且满足不同应用、不同的服务质量需求，认知无线电网络中的认知用户必须：

1）判断可用的频段，称为频谱感知。

2）选择最佳的可用频段，称为频谱判决。

3）与其他用户协商接入该频段，称为频谱共享。

4）当检测到授权用户时放弃该频段，称为频谱移动性。

正如在参考文献［5］中提到的，每个认知无线电具有认知、重配置和自组织的能力，从而能够执行所需的频谱感知、频谱判决、频谱共享和频谱移动性功能。

在参考文献［7］中概述了不同的频谱共享模型，包括共享、分层接入和动态排它使用模型。频谱管理是认知无线电网络中的一个重要功能，包括动态频谱接入/共享和定价，以满足主用户和次要用户的需求。参考文献［6］还提到 FCC 频谱策略任务组获得的频谱使用实际测试结果告诉我们：在任何特定的时间和位置，很多宝贵的频谱都是空闲的。这个矛盾告诉我们频谱稀缺是由于频谱管理策略造成的，而不是由于可使用的频段稀少。频谱利用率不足已经引起了工程、经济和管理机构开始行动起来寻找更好的频谱管理政策和技术。

5.1.1 认知无线电的能力

认知无线电技术是动态提高频谱利用率的关键技术之一。Haykin 在参考文

献［8］中定义认知无线电如下：

认知无线电是一个智能无线通信系统。它能够感知外界环境（也就是外部世界），并使用人工智能技术从环境中学习，通过实时改变某些操作参数（比如传输功率、载波频率和调制技术等），使其内部状态适应接收到的无线信号统计特性变化，以达到以下目的：任何时间、任何地点的高度可靠通信；对频谱资源的有效利用。

认知无线电网络中，认知无线电节点的能力可以根据认知无线电定义中的功能划分。认知无线电应该能够感知环境（认知能力）、分析和学习感知到的信息（自组织能力）和适应环境（重配置能力）。

我们把认知能力归纳如下：

1）频谱感知：认知无线电可以感知频谱并检测频谱空洞，也就是没有被授权用户使用的频带或者有限干扰的频带。认知无线电可以通过在频率授权方和第三方之间达成协议后共享频谱空洞。最终参与者可能基于 Ad hoc 或实时的办法协商使用频谱，而不需要几方预先达成协议。

2）位置识别：位置识别是指能够确定所处位置和其他传输机的位置的能力，然后选择在该位置合适的工作参数，例如功率和频率。在那些用于卫星下行的频段，由于只有接收机而不发射信号，定位技术可能是避免干扰最合适的办法，因为感知技术可能不能识别出临近接收机的位置。然而，这种位置识别应该根据相关的信息、系统的灵活性和总的频谱利用率，而不是固定的位置来创建系统/网络控制负载。

3）网络/系统发现：认知无线电终端要找到最佳的通信方法，首先该终端寻找周围可用的网络。可以通过直接单跳链路或多跳中继节点连接到这些网络。例如，当一个认知无线电终端想开始电话呼叫，首先它去寻找周围是否有 GSM 基站收发台或者 WiFi 的访问接入点。如果终端到基站收发台/访问接入点之间没有直接通信链路，但是可以通过其他认知无线电终端连接到 GSM 或 WiFi 网，此时该终端仍然可以发起呼叫。寻找单跳或多跳之外的接入网的能力十分重要。

4）业务发现：业务发现一般是和网络/系统发现联系在一起的。网络或系统运营商通过接入网提供服务。认知无线电终端可以找到合适的服务来满足自己的需求。

5.1.1.1 可重配置能力

认知无线电可以看成是更一般化的 SDR 技术（见第 2 章）。但是总的来说，认知无线电与 SDR 相比要具有下面的可重配置能力：

1）频率灵活性：指的是无线电改变工作频率的能力。这种能力一般与根据感知其他发射机信号或其他方法获得的信息来动态选择合适的工作频段结合起来

工作。

2）动态频率选择：指的是动态监测其他无线系统的信息，并且避免与这些系统同频工作的机制。设备用来决定何时改变工作频率或极化方向的方法包括频谱感知、地理位置监测或网络/其他设备的指示等。这种能力可以更一般化为无线通信里动态选择逻辑信道和物理信道。

3）自适应调制/编码（AMC）：该技术用于改善衰落信道下的信道容量。AMC 可以调整传输特性和波形，以增强频谱接入能力，在其他信号出现时更好地使用频谱。认知无线电应该选择出最合适的调制类型用于特定的传输系统，从而保证系统之间的互操作性。

4）传输功率控制（TPC）：设备可以通过功率控制在数据传输过程中以不同的传输功率等级动态地切换。当需要以较大功率传输时可以在允许的功率限制内传输，但是当不需要高功率时可以降低传输功率，从而允许共享功率。

5）动态系统/网络接入：认知无线电终端要接入多个运行不同协议的通信系统/网络，就必须能够自我配置，以兼容多种接入系统。因此在多无线共存环境下动态系统/网络接入有助于完全使用异构无线网络。

5.1.1.2　自组织能力

综上所述，我们知道认知无线电应该具有感知和重配置的能力。为了更智能地与终端设备通信，认知无线电还应该能根据感知和重配置功能进行自组织通信。

1）频谱/无线资源管理：需要有好的频谱管理方案来有效地在认知无线电之间管理和组织频谱空洞信息。

2）移动性和连接管理：由于认知无线电网络的异构性，路由和拓扑信息相比传统网络更加复杂。好的移动性和连接管理方案可以有助于邻居发现、检测可用因特网接入和支持垂直切换，从而有助于认知无线电选择路由和网络。

3）信任/安全管理：认知无线电网络本质上也是异构网络，而异构性（例如不同的无线接入技术，系统/网络运营商）带来了很多安全问题。因此认知无线电网络的安全运行需要引入信任机制来支持动态环境中的安全功能。

5.1.2　动态频谱接入的频谱共享模型

与目前使用的静态频谱管理策略不同，动态频谱接入（DSA）包含了改变频谱使用方法更广的内涵。在第一届 IEEE 动态频谱接入网络新领域讨论会（DyS-PAN）上提出了这个概念。如图 5-2 所示，动态频谱接入策略可以大概分为三种模型。

5.1.2.1　动态独占使用模型

这种模型是目前频谱管理政策的基本结构：频带独家授权给特定服务。主要

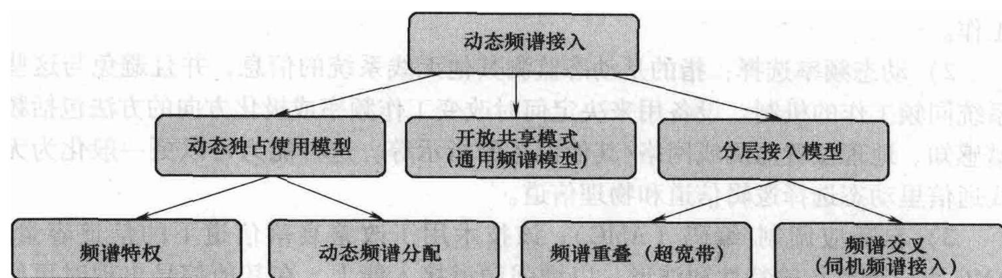


图 5-2 动态频谱接入的分类[7]

的思路是引入灵活性来提高频谱效率。在这种模型下提出了两种方法：频谱特权和动态频谱分配。前者是允许授权者出售和交易频段，并且可以自由地选择接入技术。因此经济和市场因素在更高效地使用受限频谱资源时起着重要的作用。注意，即使授权者有权租赁或共享频谱来获取利润，但是这并不是频谱管理政策委托共享的。

第二种方法，动态频谱分配是由欧盟 DRiVE 项目提出的。其目标是利用不同服务在时间和空间的业务统计特性来动态分配频谱。换句话说，在特定的区域和特定的时间，频谱分配给特定的服务独占。但是，这种动态分配比目前使用的固定分配变化得要快得多。根据独占使用模型，这些方法不能消除由于无线业务的突发特性导致的频谱白空间。

5.1.2.2 开放共享模型

这种模型也称为通用频谱，这种模型利用对等用户之间的开放共享为基础来管理频谱区域。在免使用许可的工业、科学和医疗（ISM）无线频带上的无线服务（例如 WiFi）的成功正是使用了这种模型。集中式和分布式频谱共享策略已经被提出用于解决开放共享频谱管理模型带来的技术挑战。

5.1.2.3 分层接入模型

这种模型使用主用户和次要用户的分层接入结构。基本的思路是向次要用户开放需要频谱许可的频带，同时限制对主用户的干扰。有两种在主用户和次要用户之间共享频谱的方法：频谱重叠和频谱交叉。

频谱重叠方法通过严格限制次要用户的传输功率让次要用户在主用户的噪声功率下工作。通过扩频把传输信号扩展到很宽的频带（UWB），次要用户可以特别低的传输功率实现短距离高速传输。假设在最差的情况下，主用户一直在传输，这种方法不依赖于检测和利用频谱白空间。

频谱交叉方法是 Mitola 首先基于频谱池的概念提出的，然后在 DARPA 下一代（XG）项目中提出了伺机频谱接入。与频谱重叠不同，频谱交叉不严格限制次要用户的传输功率，而是限制次要用户传输的时间和地点。允许次要用户用非

入侵的方式识别和利用本地暂时可用的频谱，从而利用空间和时域上的频谱白空间。与动态独占使用和开放共享模型相比，这种分层模型可能是与目前频谱管理政策和合法的无线系统最兼容的模型。此外，还可以同时使用频谱重叠和交叉方法来进一步提高频谱利用率。

5.1.3 伺机频谱接入：基本组成部分

我们主要讨论分层接入模型下的频谱交叉共享方法（见图5-2），也会使用伺机频谱接入（OSA）这个术语。OSA的基本组成部分包括识别频谱机会、使用频谱机会和管理政策。机会识别模块负责准确识别和智能跟踪在时间和空间上空闲的动态频带。机会利用模块则根据机会识别模块的输出来决定是否在空闲频带上传输和如何传输。管理政策定义了次要用户的基本频谱礼仪以确保与合法系统的兼容性。

OSA的整体设计目标是在保护频谱授权用户不受干扰的前提下为次要用户提供足够的利益。次要用户对传输性能的需求和主用户对干扰保护的需求之间的矛盾表明了机会识别、机会利用和管理政策几部分功能之间的交互。因此优化设计OSA需要从信号处理层面、网络互连和制定频谱管理政策等各方面跨层考虑。

5.1.4 认知无线电的网络互连

最近的研究（例如参考文献[5-7]）表明仅建立一条CR链路（从CR发射端到CR接收端）远远不够，我们需要设计开发出让认知无线电互连的方法，也称为认知无线电网络（CRN）。CRN可以部署成网络集中式、分布式、Ad hoc和网状网络结构，并且能满足授权和非授权应用的需求。同样，CRN用户可以与其他用户通过多跳的方式或者接入基站的方法通信。因此，CRN有以下三种不同的接入类型：

1）CR网络接入：CR用户可以在授权频带和非授权频带接入其网络基站。

2）CR Ad hoc接入：CR用户可以在授权频带和非授权频带与其他用户通过Ad hoc方式通信。

3）主网络接入：CR用户还可以在授权频带接入主系统基站。

如上所述，CR用户可以工作在授权频带和非授权频带。但是，CRN所需的功能取决于工作频带是在授权频带还是非授权频带。在授权频带和非授权频带，CRN的一些特性和限制有很多不同。我们可以把CR频谱应用划分为三种可能场景，如图5-3所示：图a所示为在授权频带上的CRN；图b所示为在非授权频带工作的CRN；和图c所示为同时工作在授权频带和非授权频带的CRN。

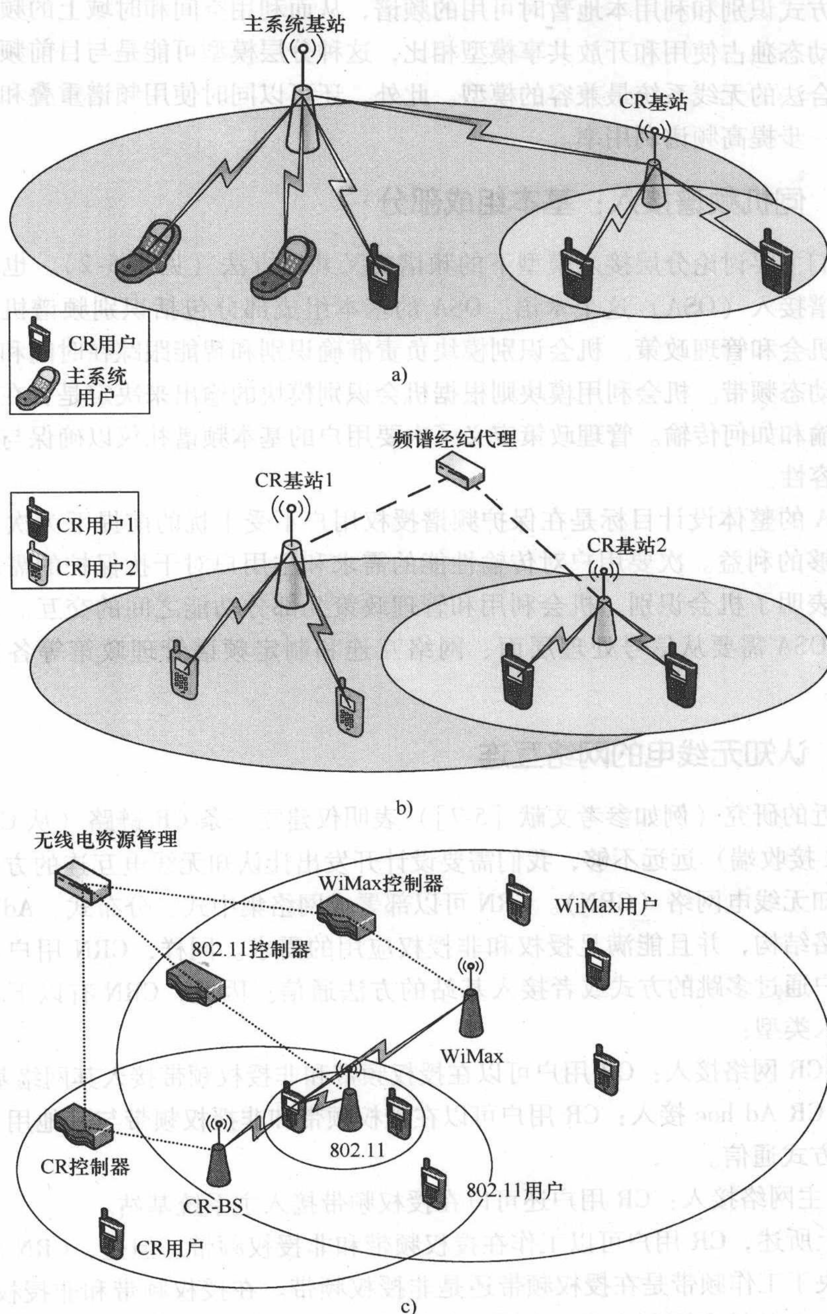


图 5-3 a）在授权频带上的 CRN；b）在非授权频带工作的 CRN；
c）同时工作在授权频带和非授权频带的 CRN

5.2 动态频谱接入的分析方法和算法

为了允许认知无线电能利用频谱空洞（频谱可用性）来进行传输，也就是CR系统要在不干扰主系统（PS）的前提下与主系统在同一频带能同时工作，我们需要考虑各种可能性，例如频谱重叠、交叉和交织。

频谱重叠方法通过强制次要用户信号功率低于频谱掩膜，从而使得次要用户/节点产生的干扰在主用户的可接受干扰功率水平之下，因此主系统和认知系统可以同时传输。一般次要系统信号是扩频信号，并且局限于短距离无线通信，例如超宽带（UWB）系统。

频谱交叉方法允许主系统和次要系统同时传输。次要用户/节点可以使用一部分功率用于次要传输，其余功率辅助主系统传输，可以看成协作中继的一种[13]。通过选择合适的功率，协作中继给主系统带来的信噪比的提高可以抵消由于次要用户干扰带来的信噪比的降低。可获得信道状态信息和成熟的编码技术（例如脏纸编码）被广泛用于减少次要用户/节点的干扰。

频谱交织方法是基于从J. Mitola最初想法得到的伺机通信。在频带上存在有暂时空闲的频带，也称为频谱空洞，这些频谱空洞没有被授权主用户占据使用。这种频谱空洞随着时间和地理位置动态出现。因此，CR必须通过物理层频谱感知持续监测频谱，然后使用某些媒体接入策略来占据频谱空洞，用于次要系统的传输，同时还要保证对主系统的节点/用户干扰最小（最好是没有干扰）。

5.2.1 开放频谱中的动态频谱接入

认知无线电的概念已经在多无线电系统共存场景中体现出在提高总体频谱利用效率方面的优点。下面是三个很好的例子：

1）蓝牙设备需要和许多免频段许可的无线设备在2.4GHz ISM（工业、科研和医学）频段上共存，例如无线局域网（也就是WiFi）设备和微波炉。在蓝牙2.0规范和IEEE 802.15.2协议中采用了自适应跳频的关键技术。自适应跳频由两个步骤组成：a）把ISM频带划分为几个子带（一般是三个信道），扫描这些子频带，并检测正在被其他无线设备使用的子频带（主用户或WiFi设备）；b）调整跳频序列来避免使用被占用的频带，从而确保蓝牙ACL和SCO链路的传输质量。

2）正在研究中的IEEE 802.11h协议考虑了在5GHz U-NII频带和ISM频带工作的情况。在某些区域，主用户的传输功率可能很高，但是传输占空比较低（例如用于气象监测的多普勒雷达）。IEEE 802.11a设备应该检测这种主用户传

输，并且利用那些主系统未占用的信道来传输自己的信号。

3）对于全球使用的超宽带通信，信号传输带宽很大，一般需要采用检测和避免（DAA）技术来检测主用户，从而避免干扰主用户的传输。

因此，在非授权频带上的传输为认知无线电概念提供了很好的示例。假设可以实现可靠的频谱感知（见第7章）。下一个技术挑战就是如何建模开放频谱无线网络的接入。图5-4给出的是在一个频带内两类无线系统共存的例子[16]。如图所示，A类系统可以使用3个宽带频率格，而B类系统可以使用9个窄带频率格。假设两类无线电设备都是用传输前监听机制（LBT，实际上也是CSMA的一类）。每个类型系统的数据到达过程服从参数为λ_i，$i=a$，b的泊松过程和服务速率参数为μ_i，$i=a$，b的指数分布，两类系统相互独立。

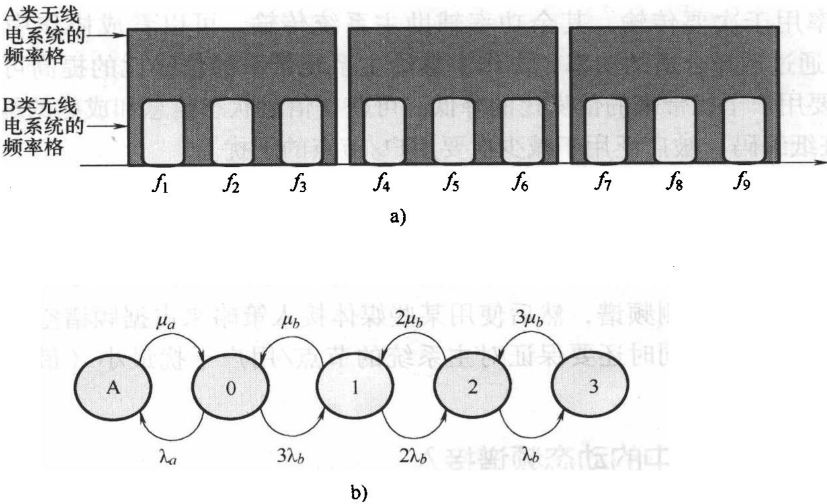


图5-4 多无线电系统共存和接入的马尔科夫链模型：a）A类和B类无线电设备的频率信道；b）连续马尔科夫链模型

这种非授权频谱接入可以建模成时间连续马尔科夫链，如图中所示定义了以下五个状态：

状态A：无线电系统A占用整个频带；

状态0：所有三个宽带频率格都是空闲；

状态1：只有一个B类系统在工作；

状态2：有两个B类系统在工作；

状态3：有三个B类系统在工作；

通过定义马尔科夫链的状态转移矩阵$\boldsymbol{\Phi}$和总的稳定状态分布概率，我们可以通过线性等式获得稳定状态分布矢量$\boldsymbol{\pi}=[\pi_A,\pi_0,\cdots,\pi_3]$：

$$\pi\Phi=0$$

$$\sum_{i=A}^{3}\pi_i = 1$$

并且

$$\Phi=\begin{bmatrix} -\mu_a & \mu_a & 0 & 0 & 0 \\ \lambda_a & -\lambda_a-3\lambda_b & 3\lambda_b & 0 & 0 \\ 0 & \mu_b & -\mu_b-2\lambda_a & 2\lambda_a & 0 \\ 0 & 0 & 2\mu_b & -2\mu_b-\lambda_b & \lambda_b \\ 0 & 0 & 0 & 3\mu_b & -3\mu_b \end{bmatrix} \quad (5\text{-}1)$$

根据这种模型可以分析其他网络特性。

5.2.2 伺机频谱接入

伺机频谱接入（OSA）是分层 DSA 的一种，当主系统用户没有占用信道时允许次要用户接入信道。我们应该在不影响主系统用户的限制前提下优化频谱利用效率。

考虑主用户根据调度接入或随机接入信道的分层接入模型，CR 用户/节点伺机接入信道传输。假设在频谱内一共有 N 条可用并行信道（编号从 0 到 $N-1$）。每条信道独立被主系统用户占据，从而可以数学建模成在上一节介绍的时间连续马尔科夫链，其中包括空闲状态（$X_i=0$）和繁忙状态（$X_i=1$）。空闲状态的保持时间服从参数为 λ_i^{-1} 的指数分布，繁忙状态的保持时间服从参数为 μ_i^{-1} 的指数分布。次要用户（也就是 CR 用户）执行分时隙协议，并在每个时隙进行下列操作[17]：

1）在每个时隙开始时感知信道。

2）如果有需要传输的业务，在这 N 个信道中选择传输的信道。

3）接收成功传输的确认信息（可能需要在认知网络中讨论这个假设，实际上，不可能除了接收确认信息外同时进行其他操作，因为伺机接入的可用时间窗口非常短）。

为了简化分析，我们假设频谱感知是周期进行的，这个假设虽然不够精确但是比较合理，因为很多无线电系统都是分时隙的，并且使用信标信号用于定时参考。每个 CR 设备连续的在协议周期的 N 个时隙感知 N 个信道。在每个时隙的开始时刻，CR 感知相应的信道（也就是在第 n 个时隙内感知第 n 个信道），并且感知持续时间要比时隙长度短得多。根据当前和过去的感知结果，CR 决定是否在 N 个信道中的某个信道传输（不局限于正在感知的时隙）或是不进行传输。

对第 k 个时隙，$I_k\equiv[kT_s,\ (k+1)\ T_s]$，感知信道 $q=k\bmod N$，其中 T_s表示

时隙周期。在 I_k 时刻周期感知之后，我们可以定义 N 维随机过程矢量 $\boldsymbol{Z}(\boldsymbol{k})=[Z_0(k),\cdots,Z_{N-1}(k)]^{\mathrm{T}}$，其中

$$Z_i(k)=\begin{cases}X_i(kT_s),i=k\bmod N\\Z_i(k-1),\text{其他}\end{cases} \tag{5-2}$$

式中，$i=0,1,\cdots,N-1$，事实上，$Z(\boldsymbol{k})$ 包括最近 N 个时隙感知的结果。如果在信道 n 感知到主系统用户，那么将更新 $Z(\boldsymbol{k})$ 的第 n 个元素。建模观测结果的马尔科夫链取决于频谱感知的有效期。令 $q\equiv k\bmod N$ 表示目前 N 个时隙周期的时隙位置。如果在第 k 个时隙感知信道 i，感知结果的有效期为 $L(i,k)=1$。一般地，

$$L(i,q)\equiv(N+q-i)\bmod N$$

然后，可以用马尔科夫链模型来描述观察的动态业务。

命题5.1：考虑 N 个并行信道，可以把业务流建模成独立两状态时间连续马尔科夫链。对信道 i，$i=0,1,\cdots,N-1$，令 λ_i^{-1} 和 μ_i^{-1} 分别表示状态0和状态1的平均保持时间，转移速率矩阵可以表示为

$$\boldsymbol{Q}_i=\begin{bmatrix}-\lambda_i & \lambda_i\\ \mu_i & -\mu_i\end{bmatrix}$$

因此，在式（5-2）中定义的过程矢量 $Z(k)$，$k=N,N+1,\cdots$是离散时间马尔科夫链。我们可以由此得到转移概率，从而可以得到数学模型的工程解。

5.2.3 伺机功率控制

IEEE 802.22 无线区域网中的主系统是电视广播站，另一种避免 CR 设备对主系统用户干扰的方法是对 CR（也就是次要用户）设备使用传输功率控制（TPC）[18]。我们分析工作在衰落信道下的 CR 系统，CR 可以收集到所有链路的信道状态信息（CSI）。所有 PS-Tx、PS-Rx、CR-Tx、CR-Rx 链路都在通信范围之内，信道增益如图 5-5 所示，信道衰减统计量用信道增益系数来表示。

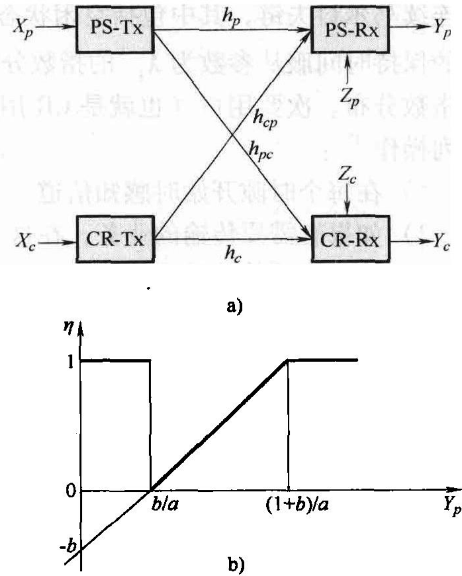


图 5-5 伺机功率控制

a）系统模型 b）选择功率控制参数

在某个时刻，PS-Rx 和 CR-Rx 接收到的信号是

$$Y_p=h_pX_p+h_{cp}X_c+Z_p$$

$$Y_c = h_c X_c + h_{pc} X_p + Z_c$$

式中，Z_p和 Z_c分别是零均值、方差为 σ_p^2和 σ_c^2的加性高斯噪声和干扰信号。当然，PS-Rx 链路并不知道 CR 的存在，它的传输功率和速率分别为 P_p和 R_p，因此我们只能假设不需要 CSI 信息的单用户检测。CR-Tx 能获得所有链路的 CSI 信息，所以 CR-Tx 可以在［0，$P_{c,\max}$］范围内改变传输功率 P_c来满足下列目标：

1）在 CR 链路上传输尽量多的信息；

2）保持主系统的中断概率不受 CR 传输的影响。

因此在没有 CR 系统时，PS-Rx 的信噪比是

$$\gamma_p = \frac{|h_p|^2 P_p}{\sigma_p^2}$$

在有 CR 系统传输的场景中，PS-Rx 的信噪比（更准确地说是信号与干扰加噪声比（SINR））为

$$\gamma_p(\eta) = \eta \frac{|h_p|^2 P_p}{|h_{cp}|^2 P_{c,\max} + \sigma_p^2}$$

CR-Rx 的 SINR 为

$$\gamma_c(\eta) = \eta \frac{|h_c|^2 P_{c,\max}}{|h_{cp}|^2 P_p + \sigma_c^2}$$

我们使用参数 $0 \leqslant \eta \leqslant 1$ 用于在［0，$P_{c,\max}$］之间的功率控制。现在的问题是如何获得所有的信道增益信息，这也是频谱感知的首要工作之一。

伺机功率控制可以在以下两种情况下考虑：

1）主链路出现中断：CR-Tx 可以全功率 $P_{c,\max}$传输。

2）主链路未出现中断：CR-Tx 要确定合适的参数 η、P_p和 R_p，从而不干扰主系统。

为了不引起主系统链路中断，应该要选择合适的 η 确保主系统链路在 PS-Rx 处的 SINR 要高于 2^{R_p-1}，也就是，CR-Tx 应该根据下式调整其传输功率

$$\eta = \min\left\{\left[\frac{|h_p|^2 P_p}{2^{R_p-1}} - \sigma_p^2\right], 1\right\}$$

同样地，对于任意给定的 h_{cp}，选择 η 的判决规则如图 5-5b 所示，其中 $a = \frac{b}{2^{R_p-1}}$，$b = \frac{\sigma_p^2}{|h_{cp}|^2 P_{c,\max}}$。这种最优功率控制策略能在不影响主系统中断概率的限制下最大化 CR 用户的速率。

5.3　认知无线电的基本限制

正如在之前章节描述的，Mitola 把 SDR 的定义推进了一步，提出可以根据周

围的无线电环境来改变网络，调制和编码参数的无线电，也称为智能无线电或认知无线电。这种无线电还可以根据周围协作节点的可用性，或无线电所处位置和频谱状况等规则来做出判断。从2000年开始，FCC已经开始初步开发次要市场，还发布了关于认知无线电技术使用的不同规则。FCC正在移除新的面向次要市场策略中不必要的管理障碍，例如：

1）频谱租赁：允许非授权用户租赁授权用户的部分频谱或者全部频谱。

2）动态频谱租赁：临时的或伺机的使用频谱，而不是长期租赁。

3）私有共有：授权方可以允许非授权用户不需要许可接入其占用的频谱，并且可选择接入费用。

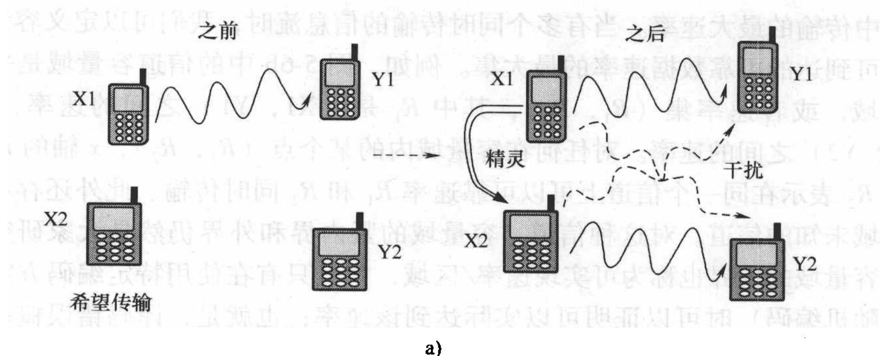
4）可中断频谱租赁：适合于需要确保当认知无线电临时占用或租赁频谱时，租赁方可以有效地收回频谱的情况。

可中断频谱租赁最好的例子是租赁美国政府或本地执法机构没有占用的频段，当应急业务需要占用时可以迅速收回频谱。可中断频谱租赁方法类似于频谱池方案。在目前FCC对伺机信道使用的提案中，认知无线电监听无线信道并且判断在时间和频域上哪些资源是未被占用的[1]。然后认知无线电在不同的时间段或不同的频段上传输信号以占据空闲频段。因此，认知无线电设备只在没有其他用户使用时才在某些时段或频段上传输。

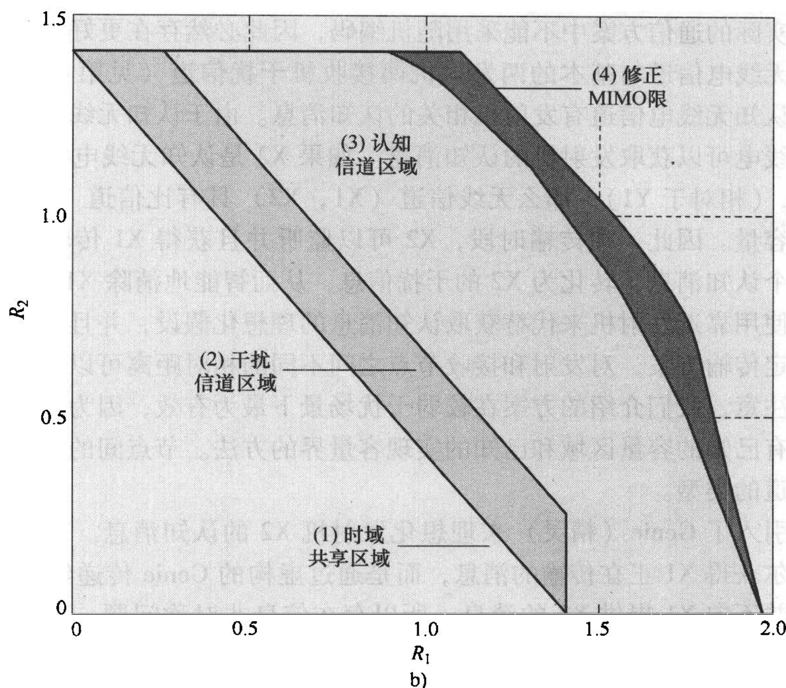
另一种更灵活、频谱效率更高的方法是允许两个用户同时在同一频率上传输。在这种方案中，一个认知无线电用户监听信道，如果监听到信道是空闲的，认知无线电可以按目前的提案进行（也就是在空闲频段上传输）。另一方面，如果另一个发射机在监听，无线电可能会进行同时通信。认知无线电不需要等待空闲信道来开始传输。这种新模型存在着下面的挑战：1）是否比时间共享频谱的效率更高？2）两个用户传输最大可达到的速率多大？与非认知无线电设备同样处理的场景相比如何？N. Devroye、P. Mitran 和 V. Tarokh 在参考文献［12］中对这些问题提出了解答。

认知无线电可以监听周围的无线信道环境，并做出判决，使用不同的方案进行编码。为了完全利用这些认知能力，首先考虑如图5-6a的一个最简单的例子，认知无线电设备可以占用信道来提高频谱利用率。假设发射机X1在无线信道中传输数据到接收机Y1，一个次要用户X2想传输数据到接收机Y2。在目前的次要频谱授权提案中，次用用户X2通过认知无线电感知其他传输用户的存在，一直等待到X1完成传输才开始传输，或者在别的频带上传输。并不强迫X2等待信道空闲，而是允许X2能够和用户X1在同一时段的同一频段上传输。信道的无线本质使得同时传输的用户之间干扰不可避免。然而，通过利用认知无线电的能力，可以证明认知无线电可以减少同时同频传输的干扰。

信息论的贡献之一是提出了信道容量。信道容量是定量的信息可以可靠地在



a)



b)

图5-6 a）认知无线电信道定义成两个发射机（X1，X2）和两个接收机（Y1，Y2）的干扰信道，其中认知无线电传输机 X2 与 X1 计划传输的消息是非因果的关系。X2 可以减少其他用户的干扰，辅助 X1 传输消息，正如我们提出的，平滑混合两个用户。b）不同的两发射机、两接收机无线信道的速率域（R_1，R_2）。区域（1）是两个独立发射机的时间共享区域。区域（2）是干扰信道的可到达域。区域（3）是认知无线电信道的可到达域。区域（4）是认知无线电信道容量的外界。所有的仿真都是基于 AWGN 信道，发射机功率为6，噪声功率为1。干扰信道的交叉参数为0.55和0.55（见参考文献［12］中的图3）

信道中传输的最大速率。当有多个同时传输的信息流时，我们可以定义容量域为同时可到达的可靠数据速率的最大集。例如，图 5-6b 中的信道容量域是一个二维区域，或者速率集（R_1，R_2），其中 R_1 是（X1，Y1）之间的速率，R_2 是（X2，Y2）之间的速率。对任何在容量域内的某个点（R_1，R_2），x 轴的 R_1 和 y 轴的 R_2 表示在同一个信道上可以可靠速率 R_1 和 R_2 同时传输。此外还存在很多容量域未知的信道。对这种信道，容量域的紧内界和外界仍然是大家研究的目标。容量域的内界也称为可实现速率/区域，并且只有在使用特定编码方案（一般是随机编码）时可以证明可以实际达到该速率；也就是，译码错误概率随着分组大小的增大，逐渐减少到 0。注意这就保证了存在着可以这些速率可靠传输的方案。实际的通信方案中不能采用随机编码，因此必然存在更好的编码方案。

认知无线电信道与基本的两发射机两接收机干扰信道（见第 4 章）不同的地方在于认知无线电信道有发射机相关的认知消息。由于认知无线电的特性，所以认知无线电可以获取发射机的认知消息。如果 X2 是认知无线电设备，并且位置靠近 X1（相对于 Y1），那么无线信道（X1，X2）具有比信道（X1，Y1）更大的信道容量。因此，在传输时段，X2 可以监听并且获得 X1 传输的消息，然后利用这个认知消息，转化为 X2 的干扰信息，从而智能地消除 X1 带来的干扰。虽然我们使用靠近发射机来代替获取认知消息的理想化假设，并且提出了这种场景下的特定传输方案，对发射和接收节点之间不同的相对距离可以利用不同的方案。必须注意，我们介绍的方案在较弱干扰场景下最为有效，因为强干扰和特强干扰信道有已知的容量区域和已知的实现容量界的方法。节点间的相对位置决定了干扰信道的类型。

我们引入了 Genie（精灵）来理想化发射机 X2 的认知消息。也就是我们假设不是偶尔获得 X1 正在传输的消息，而是通过虚构的 Genie 传递给 X2 该消息。注意我们并不向 X1 提供 X2 的消息，所以存在信息非对称问题。这样的理想化假设可以提供任何实际场景的容量上界，并且可以让在这种场景下使用基础技术更好地解决这个问题。

一个简单的认知无线电信道可以看成是一个两发射机两接收机的经典信息论干扰信道，并且其中发射机 2（认知无线电）获得或通过 Genie 获得发射机 1 要传输的消息。这种场景如图 5.6a 所示。认知无线电可以同时在同一个信道上传输，而不是像在传统认知无线电信道协议中那样必须等待有空闲信道时才能传输。目前已经发现了速率（R_1，R_2）的可到达域，也就是 X1 和 X2 可以同时以速率 R_1 和 R_2 传输数据到 Y1 和 Y2，其中把脏纸编码（或 Gel Fand-Pinsker）技术融入到 Han 和 Kobayashi 构建的干扰信道可到达域以及中继信道可到达域。当 X2 有 X1 将传输的先验信息时，或者是有 X2 处的干扰信息时，可以考虑以下两种可能的策略：

1）该用户可以自私地尝试抑制干扰。通过脏纸编码技术可以实现干扰抑制。在这种情况下，X2 分离出要传输到 Y2 的独立信息。这种策略将带来在图 5-6b 的认知信道域中较大 R_2、较低 R_1 的点。

2）该用户可以无私地充当增强用户 X1 信号的中继。虽然这种方案不允许 X2 传输该用户自己的独立信息，但是从公平的角度看却十分正确。因为 X2 侵占了 X1 所拥有的频谱，所以 X1 应该获得一些利益补偿才能保持公平。这种策略将带来在图 5-6b 的认知信道域中较高 R_1、较低 R_2 的点。

图 5-6b 中的认知信道域描述的是在出现加性高斯白噪声时的可到达域。在图中可以看到有四个区域：

时间共享区域（1）显示了用户 X1 和 X2 在时域上共享无线信道的容量域。这个区域内的点都是让 X1 在一部分时间段内传输，X2 在这些时间内不传输，而在另外的时间段 X2 传输，X1 不传输。这些点应该遵从次要频谱租赁中目前使用的协议。

干扰信道区域（2）对应经典的信息论干扰信道的可到达容量域。在这个区域，发射机独立编码，接收机之间没有互相的消息。

认知信道区域（3）是本章描述的可到达容量域。在这种情况下 X2 通过 Genie接收到 X1 的消息，X2 使用结合了干扰消除的编码方案来中继 X1 的消息。正如预期的，这个容量域是凸光滑的。可以把凸性看成是时间共享的结果：如果任意两个（或多个）方案可以达到一定的速率，那么把这些方案在时间上共享，就能达到任何凸的合并速率。这个区域光滑是因为我们的方案实际上包括编码级的功率共享，因此边界更加圆滑。可以看到使用这种方案不仅仅是 X2 用户能获得额外认知消息，两个用户都能获得好处。因为自私策略提高了 R_2 速率，而无私策略可以提高 R_1 速率；因此，合并这两种策略可以给两个用户都带来好处。认知无线电 X2 的出现可以给 X1 带来好处，这在实际应用中具有重大的意义。这也为引入这种方案提出另一个优点。

修正 MIMO 区域界（4）是 2×2 多入多出高斯广播信道的容量区域的外界，我们把传输协方差矩阵限制在下面的形式

$$\begin{bmatrix} P_1 & C \\ C & P_2 \end{bmatrix}$$

从而更接近我们的限制，与（X2，Y2）在出现 X1 带来干扰时 R_2 的容量限交叉。

5.4 认知无线电网络的数学模型

认知无线电的研究是一个正在兴起的研究领域，其中有不同的方法处理不同

的问题。研究者通过各种不同的途径建模问题。因此，我们在本节提出了一些十分重要的认知无线电（CR）和认知无线网络（CRN）相关的数学模型。在本节，CRN 可以狭义定义成把一些 CR 聚集在一起执行同样的网络协议，也可以广义定义成把 PS 节点和 CR 节点聚集起来在现有的主系统网络的基础上执行同样的 CR 网络协议集。

5.4.1 CR 链路模型

假设一个主授权通信系统（或者一对接收机和发射机）正在工作，一个认知无线电也就是频谱的次要用户，利用信道状态，寻找利用空闲频谱进行通信的可能性。信道可以建模成一个具有两种可能状态的 Elliot-Gilbert 信道[11]：1）存在主用户（也就是不允许任何次要用户传输的状态）；2）不存在主用户（允许次要用户传输的状态）。Stefan Geirhofer、Lang Tong 和 Brian M. Sadler 在参考文献［19］中测量了 WLAN 系统，从而统计地建模了在白空间认知无线电传输的半马尔科夫模型，如图 5-7 所示。

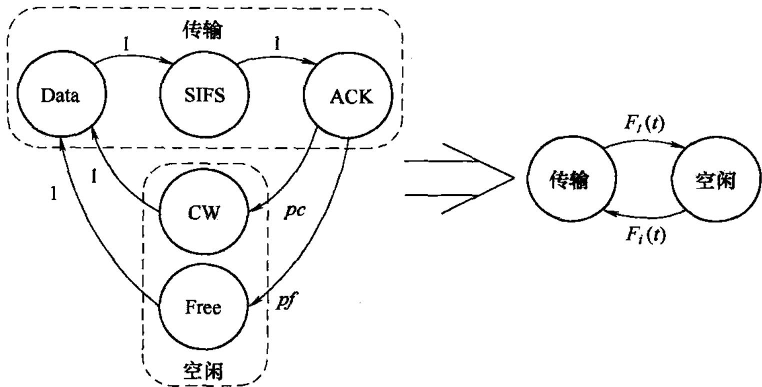
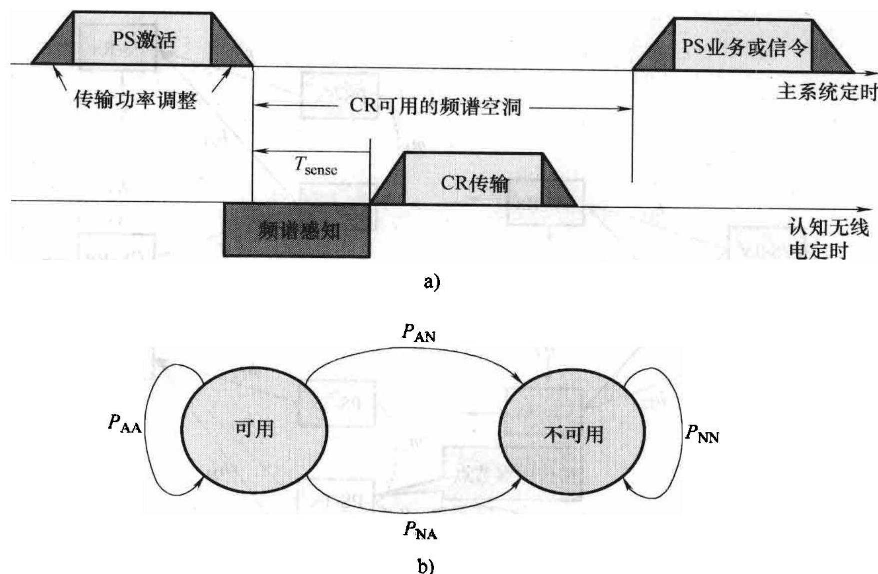


图 5-7 WLAN 的白空间模型（参考文献［19］中的图 3）

假设有一个 Genie 用于检测 PS 和 CR 的运行状态，CR 必须使用频谱空洞窗来完成数据包的传输。假设这种频谱空洞窗周期用 T_{window}表示，很明显

$$T_{\text{window}} \geqslant T_{\text{sense}} + T_{\text{CR-Transmission}} + T_{\text{ramp-up}} + T_{\text{ramp-down}} \tag{5-3}$$

式中，T_{sense}表示要确保 CR 传输机会和获取相应通信参数所需要的最小感知时间；$T_{\text{CR-Transmission}}$是 CR 数据包的传输时间；$T_{\text{ramp-up/down}}$表示传输的突变（上或下）时期。这个等式忽略了发射-接收对的传播时延和处理时延，传播和处理时延可以看成是突变上/下时间的一部分。频谱空洞（可用频谱）的最大持续时间可以认为是信标信号的持续时间。当信标信号间隔固定时，图 5-7 所示的连续时间马尔科夫链可以简化为如图 5-8b 所示的离散时间马尔科夫链。



a)

b)

图 5-8 CR 链路可用性模型

a）CR 传输机会窗口 b）链路可用性内嵌马尔科夫链的状态转移

5.4.2 重叠 CR 系统

在 5.2 节我们把 CR（或者更一般化的次要无线电）分为基于已知干扰假设的重叠模型和基于干扰避免概念的交叉模型（实际上也是交织模型）。

5.4.2.1 具有已知干扰的重叠模型

考虑 CR 在 PS 的频段上通信，如图 5-9 的左侧所示，CR-Tx 想发送消息到 CR-Rx，并且 CR-Tx 具有 PS-Tx 传输消息 m_1 的先验信息。发射机和接收机知道所有的信道增益，因此 CR-Tx 有两个策略：

自私方法：CR-Tx 采取贪婪策略，使用所有可用功率来传输消息到 CR-Rx，并且使用 PS 的先验信息利用脏纸编码技术来有效地抑制 CR-Rx 处的干扰[14]。这就是为何 h_{12} 在图 5-9a 未出现的原因，而 PS-Rx 受到 CR-Tx 通过 h_{12} 带来的额外干扰。为了使得 CR 概念能够工作，我们需要限制从 CR-Tx 到 PS-Rx 的干扰，因此在 CR 用户间（或 CR 网络）具有最大吞吐量的上限。

无私方法：CR-Tx 分配部分功率来协助 PS-Tx 传输消息 m_1 到 PS-Rx。剩下部分功率用于传输 CR-Tx 的消息。

5.4.2.2 干扰避免的交织模型

假设在重叠模型中认知设备具有所有链路的先验信息，将不会存在隐藏终端的问题（见第 7 章）。先进的编码技术可以用来实现 CR 性能上限。但是实

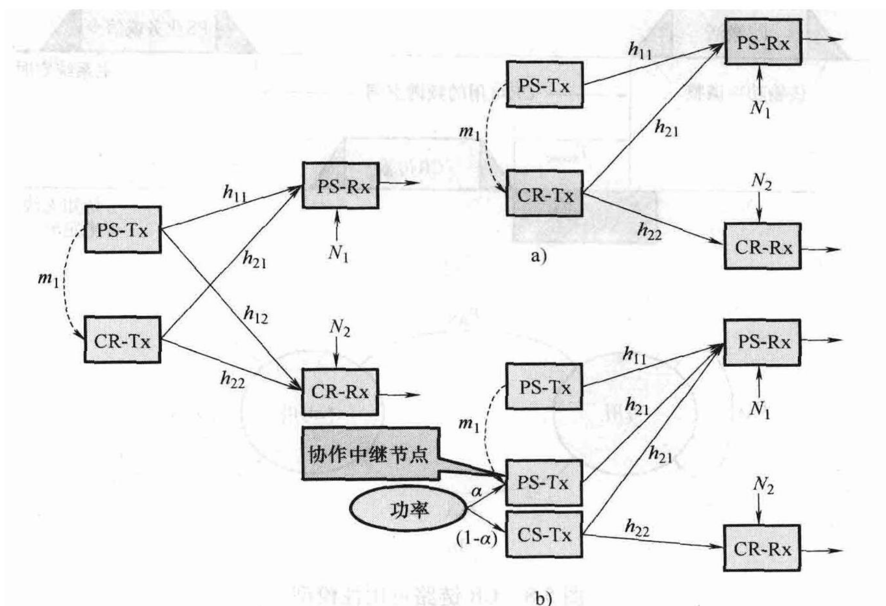


图 5-9 交叉 CR 模型

a）自私方法 b）无私方法

际上很难得到干扰信息的先验信息。因为 CR 必须面对干扰，我们可以构建能容忍干扰的接收机或者 CR-Tx 使用未被占用的 PS 频谱来避免干扰的机制。图 5-10 介绍的是 2-切换交织模型。CR-Tx 和 CR-Rx 由可能的 CR 链路和一些 PS 节点组成。虚线区域是 CR-Tx 和 CR-Rx 的感知区域，在感知区域内 CR-Tx 或 CR-Rx 可以检测到 PS 节点的活动。例如，在位置 A 和位置 B 的 PS 节点可以被 CR-Tx 检测到，而在位置 B 和 C 的 PS 节点可以被 CR-Rx 检测到。为了避免 CR 链路内的干扰，如图 5-10a 所示，在建立 CR 链路之前，我们需要在 CR-Tx 和 CR-Rx 都进行频谱感知，以检测该段频谱的可用性。此外，CR 链路的可用性检测和CRN 的运行应该是动态和分布式的。2-切换模型如图 5-10b 所示。随机变量X 和Y 表示 CR 链路的输入-输出关系，Z 为加性干扰和噪声。两个切换函数 S_t和 S_r实际上分别是表示根据 CR-Tx 和 CR-Rx 感知到的主系统用户活动的标识函数。

$$S_t=\begin{cases}1，\text{如果没有检测到激活的 PS 用户就切换到 CR 链路传输}\\0，\text{如果 CR-Tx 检测到 PS 节点活动，切换打开}\end{cases}$$

$$S_r=\begin{cases}1，\text{如果没有检测到激活的 PS 用户就切换到 CR 链路传输}\\0，\text{如果 CR-Rx 检测到 PS 节点活动，切换打开}\end{cases}$$

因此，根据图 5-10 中的网络拓扑，整个 CR 链路的输入/输出关系可以用下

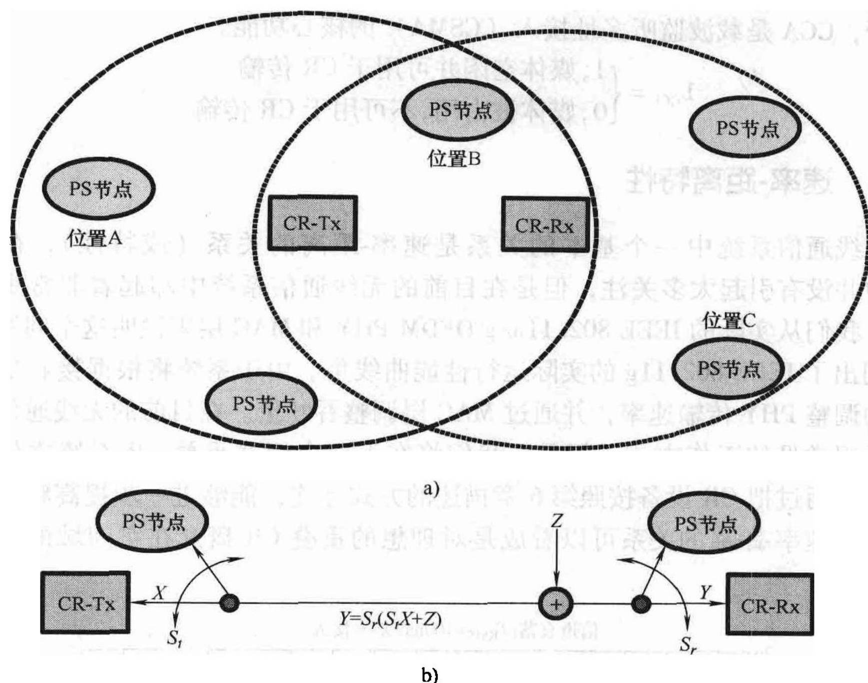


图 5-10　交织 CR 模型

a）局域频谱活动　b）交织数学模型

式表示：

$$Y = S_r(S_tX + Z)$$

一般来说，S_t和 S_r是独立工作，它们之间没有消息交互。测量 S_t和 S_r可能是相关的，并且这种相关性证明了认知无线网络的分布式特性。从这个角度来看，伺机认知无线电可以看成是具有部分边带信息的通信。我们还可以很容易地看到频谱感知对于确定建立 CR 链路的重要性，在第 7 章将详细介绍频谱感知技术，尤其是协作式频谱感知。这种模型有利于设计 CRN，因为可以直接应用在动态频谱接入（DSA）的媒体接入控制中。伺机信道选择就是一个很好的例子，也包括网络层功能。

对于干扰避免，我们可以定义媒体占用标识符（MOI）用于标识无线信道是否可以用于建立 CR 链路：

$$1_{MOI} = \begin{cases} 1，媒体被占用，因此不能用于 CR 传输 \\ 0，媒体未被占用，可以用于 CR 传输 \end{cases}$$

MOI 可以是时间、频率和位置的函数，并且可以表示物理链路甚至逻辑链路的可用性。我们还可以通过评估信道来定义另一个标识，空闲信道标识（CCI）。在 IEEE 802.11 MAC 中使用的空闲信道评估（CCA）就是 CCI 的一个实

际例子，CCA 是载波监听多址接入（CSMA）的核心功能。

$$1_{\text{CCI}}=\begin{cases}1, \text{媒体空闲并可用于 CR 传输}\\0, \text{媒体被占用，不可用于 CR 传输}\end{cases}$$

5.4.3 速率-距离特性

无线通信系统中一个基本的关系是速率-距离的关系（或特性），在通信理论中并没有引起太多关注，但是在目前的无线通信系统中却起着非常重要的作用。我们从实际的 IEEE 802.11a/g OFDM PHY 和 MAC 层来说明这个问题。图 5-11 列出了 IEEE 802.11g 的实际运行性能曲线集。由于系统将根据接收功率等级自动调整 PHY 传输速率，并通过 MAC 层调整吞吐量。在目前的无线通信系统中这是很常见的工作方式。但是，我们将在下面介绍在重叠 CR 系统之外的概念，并且通过把 CR 设备按照第 6 章描述的方式互连，能够进一步提高频谱整体利用率，速率-距离的关系可以看成是对理想的重叠 CR 概念在空间域的进一步扩展。

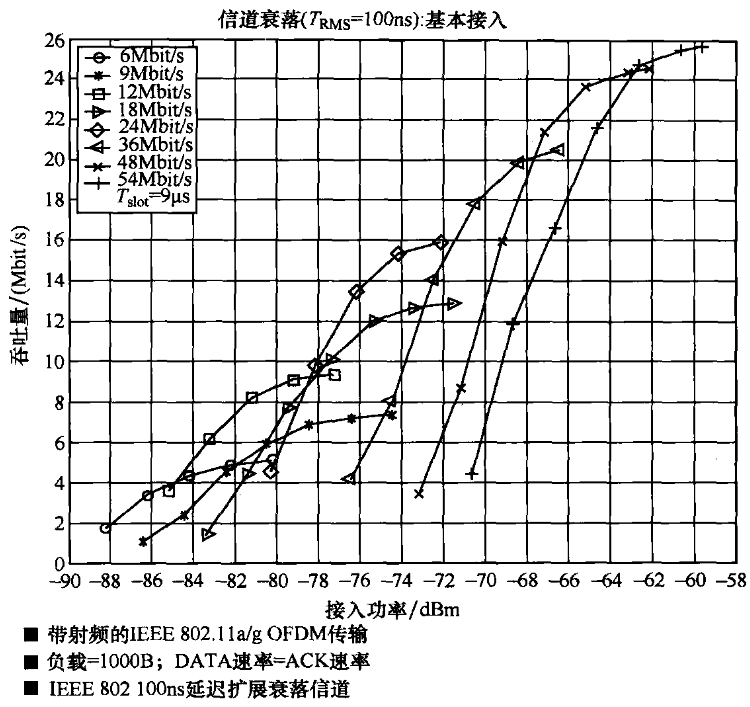


图 5-11 接收功率与 IEEE 802.11a/g OFDM PHY 数据速率（以及 MAC 吞吐量）的关系

如果考虑发射机和接收机之间的传播距离与相对的接收功率，我们可以建立起无线通信新的模型，称之为无线通信的速率-距离特性。我们还可以把这种“距离”当做测量接收信号功率的方法，而不是欧氏距离或传播距离，包括网络运行的传播因素和跨层设计。因此，距离测量 D 包括接收信号功率传播欧氏距离 D 在一定的大尺度衰落下的任何可能位置点。图 5-12a 表示了速率-距离特性，图中系统包括两种传输速率。图 5-12b 给出了次要用户对在原点处的主系统用户的最大可容忍干扰。一般认为低速率传输更容易受到干扰的影响。因此，正如图 5-12c 的结果，可以在不影响主系统用户的前提下调节次要用户的传输速率/功率。因此，在认知无线电感知到潜在的传输机会时，认知无线电的传输速率（以及功率）是由下面的速率-距离状况决定的：

1）用速率-功率分配表示的衰落信道容量；

2）共存系统的干扰水平；

3）主系统用户最大可容忍的干扰功率；

4）在主用户设备和次要用户设备之间有效的“距离”关系。

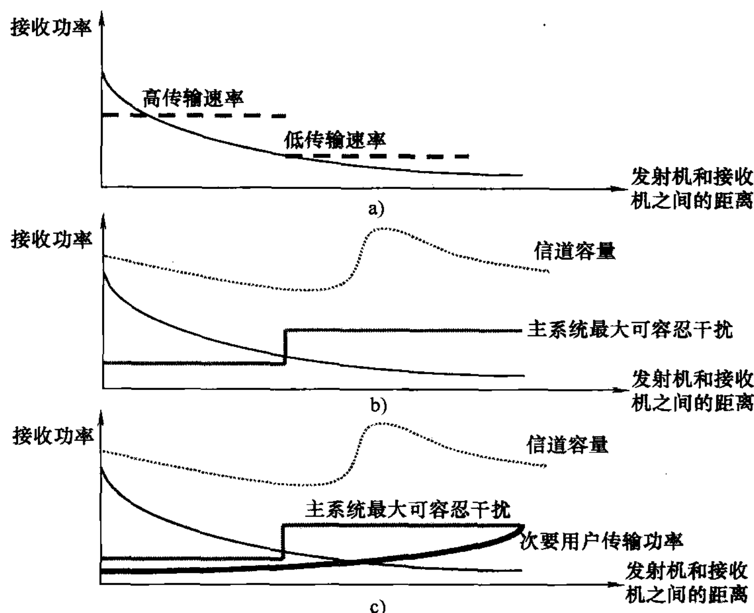


图 5-12　认知无线电的速率-距离特性

图 5-12 实际上描述的是认知无线电通信最坏的场景，图 5-13 给出的是更实际的速率-距离特性。主系统的基站和移动台在进行通信。由于主系统的有效距离有限，所以选择了低速率传输。在小区的边缘（相对基站），有两个认知无线电设备在主系统的低功率干扰的影响下尝试建立链路。如图 5-13 所示，这两个

认知无线电之间可能在不影响主系统的前提下建立起高速率通信链路，并且主系统节点可以容忍次要认知无线电带来的干扰。

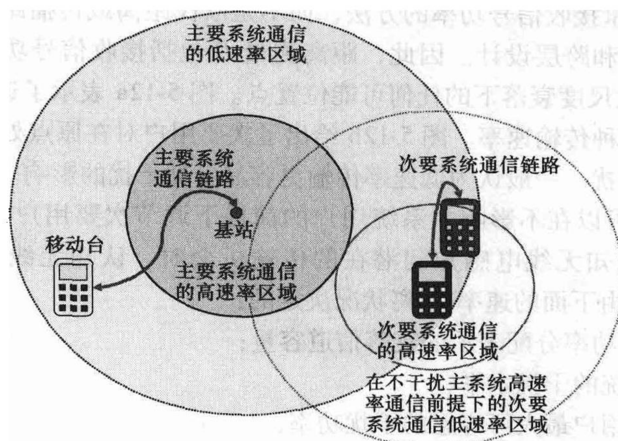


图 5-13 共存主系统/次要系统通信的速率-距离特性

事实上，这种场景下可以使用多用户检测（MUD）来减少对认知无线电的同信道干扰，因为认知无线电知道主系统用户的通信状态。从初始同步到用户识别，可以确定所有这些状态，并不局限于 CDMA 通信系统。在 OFDM 通信系统中，可以使用 MUD 来抵消同信道干扰，并且不需要主系统用户的先验信息。使用随机集理论和 MUD 技术，我们还可以进一步同时识别和译码从未知数量的用户传输来的信号。基于自适应调制和编码（AMC）技术，我们可以归纳出在没有 MUD 时，次要认知无线电的速率-功率特性的数学条件。

参考文献

[1] I.F. Akyilidiz, W.-Y. Lee, M.C. Vuran, S. Mohanty, 'A Survey of Spectrum Management in Cognitive Radio Networks', Georgia Institute of Technology.
[2] D. Cabric, I.D. O'Donnel, M. S.-W. Chen, R.W. Brodersen, 'Spectrum Sharing Radios', *IEEE Circuits and System Magazine*, 2Q 2006, pp. 30–45.
[3] NTIA, 'U.S. Frequency Allocation', [online] http://www.ntia.doc.gov/osmhome/allochrt.pdf
[4] First Report and Order, Federal Communication Commission Std. FCC 02-48, Feb. 2002.
[5] K.-C. Chen, Y.-J. Peng, N. Prasad, Y.-C. Liang, S. Sun, 'Cognitive Radio Network Architecture: Part I – General Structure', ACM ICUMIC, Seoul, 2008.
[6] K.-C. Chen, Y.-J. Peng, N. Prasad, Y.-C. Liang, S. Sun, 'Cognitive Radio Network Architecture: Part II – Trusted Network Layer Structure', ACM ICUMIC, Seoul, 2008.
[7] Q. Zhao, B.M. Sadler, 'A Survey of Dynamic Spectrum Access: Signal Processing, Networking, and Regulatory Policy', *IEEE Signal Processing Magazine*, **24**(3), 2007, 79–89.
[8] S. Haykin, 'Cognitive Radio: Brain-empowered Wireless Communications', *IEEE Journal on Selected Areas in Communications*, **23**(2), 2005, 201–220.
[9] T. Han, K. Kobayashi, Mathematics of Information and Coding, AMS 2002.
[10] A. Al-Fuqaha, B. Khan, A. Rayes, M.Guizani, O. Awwad, G.B. Brahim, 'Opportunistic Channel Selection Strategy

for Better QoS in Cooperative Networks with Cognitive Radio Capabilities', *IEEE Journal on Selected Areas in Communications*, **26**(1), 2008, 156–167.

[11] H.S. Wang, N. Moayeri, 'Finite-State Markov *Channel – A Useful Model for Radio Communication Channels*', *IEEE Tr. on Vehicular Technology*, **44**(1), Feb. 1995.

[12] N. Devroye, P. Mitran, V. Tarokh, 'Limits on Communications in a Cognitive Radio Channel', *IEEE Communications Magazine*, June 2006, 44–49.

[13] S. Srinivasa, S.A. Jafar, 'The Throughput Potential of Cognitive Radio: A Theoretical Perspective', *IEEE Communications Magazine*, May 2007, 73–79.

[14] N. Devroye, P. Mitran, V. Tarokh, 'Achievable Rates in Cognitive Radio Channels', *IEEE Transactions on Information Theory*, **52**, May 2006, 1813–1827.

[15] S.A. Jafar, S. Srinivasa, 'Capacity Limits of Cognitive Radio with Distributed and Dynamic Spectral Activities', *IEEE Journal on Selected Areas in Communications*, **25**, April 2007, 529–537.

[16] Y. Xing, R. Chandramouli, S. Mangold, S. Shankar, 'Dynamic Spectrum Access in Open Spectrum Wireless Networks', *IEEE Journal on Selected Areas in Communications*, **24**(3), 2006, 626–637.

[17] Q. Zhao, S. Geirhofer, L. Tong, B.M. Sadler, 'Opportunistic Spectrum Access via Periodic Channel Sensing', *IEEE Transactions on Signal Processing*, **56**(2), 2008, 785–796.

[18] Y. Chen, G. Yu, H.H. Chen, P. Qiu, 'On Cognitive Radio Networks with Opportunistic Power Control Strategies in Fading Channels', *IEEE Transactions on Wireless Communications*, **7**(7), 2008, 2752–2761.

[19] S. Geirhofer, L. Tong, B.M. Sadler, 'Dynamic Spectrum Access in the Time Domain: Modeling and Exploiting White Space', *IEEE Communications Magazine*, May 2007, 66–72.

[20] E Biglieri, M Lops, 'Multiuser Detection in a Dynamic Environment - Part I: User Identification and Data Detection', *IEEE Transactions on Information Theory*, **53**(9), September 2007, 3158–3170.

第6章 认知无线电网络

正如我们在第5章中描述的，传统的认知无线电致力于通过动态频谱接入来利用频谱空洞，从而达到提高频谱效率的目的。通过推广协作网络的概念，利用认知无线电终端自适应连接多个共存无线电系统的能力，我们可以构造一个认知无线电网络（CRN），其中不同的系统可以相互连接、相互协作，不论网络节点是属于主系统（PS）还是次要系统。认知无线电终端必须能够以协作中继的方式，利用现存的主系统和协作的认知无线电终端，这是对认知无线网络的一个重要推广（请参见图6-1中的示例）。我们把这样的协作式中继网络称为认知无线电中继网络（Cognitive Radio Relay Network，CRRN）。因此，CRRN包括主系统网络和协作认知无线电节点，以及利用该网络传输数据的认知无线电终端。请注意，这里的协作中继节点的目的是创建整个网络的聚合带宽，而不是一般的协作通信中创建分集。

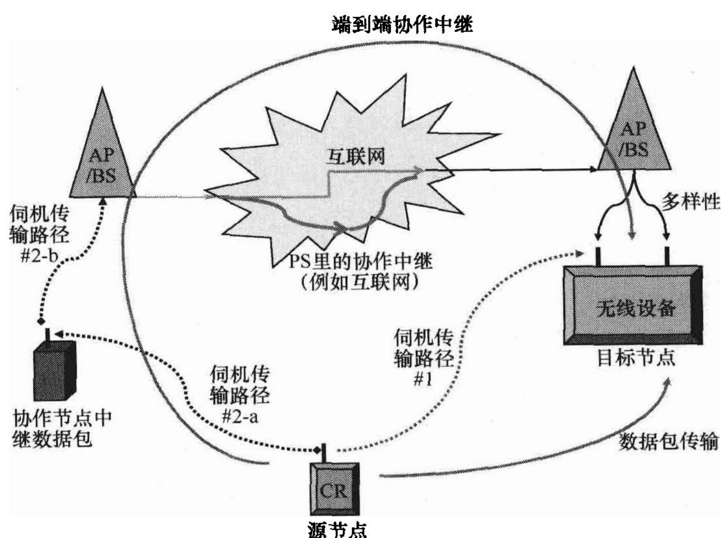


图6-1 协同中继节点带来更多网络聚合带宽

换言之，认知无线电并非只是利用频谱空洞的链路层技术。认知无线电可以和协作中继（即协作式网络）一起工作，通过利用主系统来构造CRRN。进一步地说，来自认知无线电源节点的数据包或者消息，可以通过

多跳协作式中继网络，将共存的主系统节点以及其他认知无线电节点当作中继节点，最终达到认知无线电的目标节点。本书中 CRN 的概念即是针对这样的场景的。

6.1 认知无线电中继网络的网络编码

在继续介绍 CRN 的全面功能之前，我们先集中讨论 CRRN 是否可以为网络效率带来好处，这也意味着给定频谱带宽后真正的频谱效率（而非物理层的频谱效率）。为了发掘这一基本内涵，Huang 和 Chen 在参考文献［2］中采用了信息论中的网络编码来研究 CRRN。而 Geng 在参考文献［12］中将网络编码应用于只包括 Ad hoc CR 节点的认知无线电网络。

我们用图 6-2 来说明 CRRN。主系统源用户通过主系统网络（包括中继节点，也即类似于任意数据网络的中继网络）来向目的节点传输数据包。因此主系统源节点和认知无线电节点的数据流在这个 CRRN 里共存。既然认知无线电网络的干扰会造成主系统的网络容量下降，那么我们分析认知无线电的基本行为来学习如何避免对主用户的干扰。这样，我们就可以通过最大化 CRRN 的网络容量，而且不造成对主系统业务的干扰。

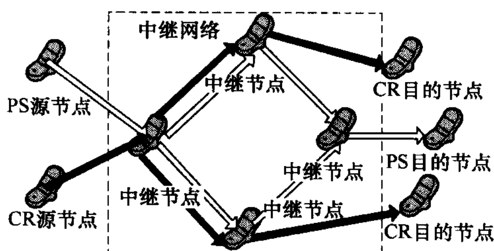


图 6-2　认知无线电中继网络

以下，我们通过分析 CRRN 中主系统最小割容量的变化来分析认知无线电和主系统之间的干扰。最小割容量是通过网络编码理论得出的整个网络的容量。我们在分析中考虑叠加编码，因为 CRRN 可能会采用译码转发的协作方式，即中继网络中的节点先对认知无线电网络的信息译码，然后将认知无线电的信息叠加在主系统的信息上。我们在译码转发的方式下分析，在不同的 CRRN 拓扑下，认知无线电的干扰是否可以避免，或者可以得到抑制。尽管 CRRN 有很多种场景，我们可以只考虑以下四种情况，作为 CRRN 的基本构成：

1）单跳中继网络（见图 6-3a）

2）汇接中继网络（见图 6-3b）

3）协作式中继网络（见图 6-4a）

4）并行协作式中继网络（见图 6-4b）

这些网络拓扑都应用在有网络编码的协作式通信中。这些关于受限或者可避

免的干扰的分析可以推广到任意协作式的 CRRN 拓扑中。根据认知无线电规避主系统的本质，我们集中研究可容忍干扰的分析。首先，我们推导使得认知无线电可以避免对主系统干扰的条件。其次，我们把认知无线电和主系统限制在单播，以简化对最大容量的分析。在保证避免对主系统干扰的前提下，我们推导出认知

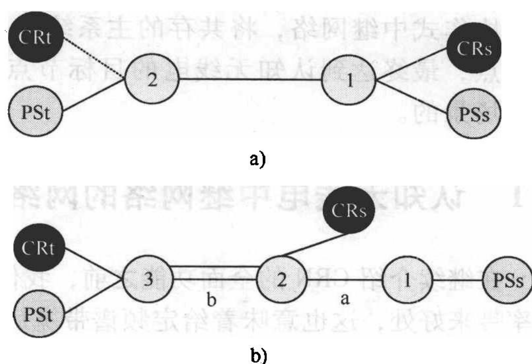
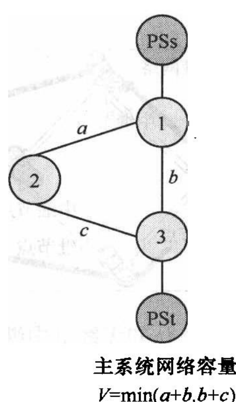
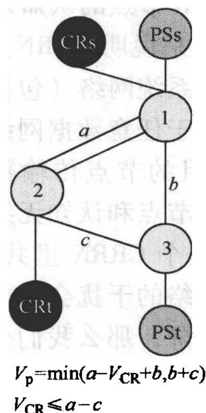


图 6-3 a）单跳中继；b）汇接中继

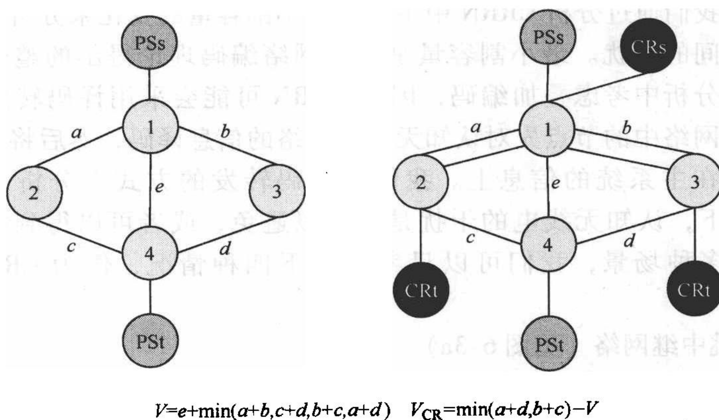


主系统网络容量
$V=\min(a+b,b+c)$



$V_{\mathrm{P}}=\min(a-V_{\mathrm{CR}}+b,b+c)$
$V_{\mathrm{CR}}\leqslant a-c$

a)



$V=e+\min(a+b,c+d,b+c,a+d)$ $V_{\mathrm{CR}}=\min(a+d,b+c)-V$

b)

图 6-4 a）协作式中继；b）并行协作式中继

无线电的最大网络容量，并将实现最大网络容量的链接分配问题公式化为多商品流问题。然后我们放宽对单播的限制，假设主系统是多播，而认知无线电是单播。我们证明在这种情况下最大化认知无线电网络容量的链接容量分配问题成为另外一个线性规划问题。因此，我们可以判决是否任意的网络都可以无干扰地被认知无线电使用，而且可以在 CRRN 中最大化认知无线电的网络容量。最后，我们也分析不可避免的干扰以及限制干扰的条件。基于这些步骤，我们可以开发出一套操作准则，以检验既可用于提高网络流量、又不干扰主系统的“机会”。

6.1.1　系统模型

因此，我们开发如图 6-2 所示的系统模型来研究 CRRN 的网络容量。

6.1.1.1　关于 CRRN 的假设

我们做如下假设，以集中分析 CRRN 中的干扰，并且简化分析过程：

1）在我们分析的 CRRN 中，只有一个主系统源节点、一个认知无线电源节点、一个或多个主系统目的节点、一个或者多个认知无线电目的节点。

2）认知无线电源节点并不直接传输到目的节点。所以，我们只考虑认知无线电被中继网络中继的数据流。

3）CRRN 中的链接一般都是单向的，因为认知无线电的链接只是机会式地存在一段时间。所以，整个网络可建模为有向图。

4）在认知无线电节点和中继网络节点之间建立新的链接，而并不改变其他链接的信道容量。

5）中继网络中的每一个边至少包含在一条从主系统源到主系统目的的路径里。换言之，中继网络里的每一条链接都要中继传输来自主系统源的数据。

我们用有向图 $G=(V,\ E)$ 来表示网络，其中不包括利用中继网络的认知无线电终端。V 是 G 中的节点集，而 E 是其中的边，容量矩阵 $\boldsymbol{R}=[R_{ij}]$ 中的元素对应于链路 $(i,\ j)\in E$。类似地，我们用有向图 $G''=(V'',\ E'')$ 来表示 CRRN，其中容量矩阵 $\boldsymbol{R}'=[R'_{ij}]$ 的元素对应于链路 $(i,\ j)\in E'$。因此，这两个图有如下的关系：

1）$V'=V\cup S_{CR}\cup t_{CR}$，$E'=E\cup E_{CR}-V$

2）若 $(i,\ j)\in E$，则 $R'_{ij}=R_{ij}$

式中，s_{CR}、t_{CR}分别是认知无线电源节点和目的节点的集合。E_{CR-V}是认知无线电节点和中继网络节点的链路。在以下各节里，我们把 $G=(V,\ E)$ 称为原始网络，而将 $G''=(V'',\ E'')$ 称为 CRRN。

6.1.1.2　译码转发协作策略：叠加编码

基于网络编码理论模型，我们定义网络编码的以下元素：

1. 消息集

$$\Omega_{PS}=\{1, \cdots, [2^{nh_{PS}}]\}, \Omega_{CR}=\{1, \cdots, [2^{nh_{CR}}]\}$$

2. 各链路上的编码函数

对于不对认知无线电的消息进行中继的链路：

$$f_{ij}: \Pi_{(i',i)\in E'}A_{i'i}\rightarrow A_{ij} \tag{6-1}$$

对于对认知无线电和主系统的消息进行中继的链路：

$$f_{ij,PS}: \Pi_{(i',i)\in E'}A_{i'i,PS}\rightarrow A_{ij,PS} \tag{6-2}$$

$$f_{ij,CR}: \Pi_{(i',i)\in E'}A_{i'i,CR}\rightarrow A_{ij,CR} \tag{6-3}$$

对于将节点 i 连接到目的节点的链路：

$$q_i: \Pi_{(i',j)\in E'}A_{i'j}\rightarrow A_i \tag{6-4}$$

3. 目的节点的译码：

对于主系统：

$$q_{PS}: \Pi_i A_i\rightarrow \Omega_{PS} \tag{6-5}$$

对于认知无线电：

$$q_{CR(i)}: \Pi_i A_i\rightarrow \Omega_{CR} \tag{6-6}$$

对应于主系统和认知无线电的消息集合为 Ω_{PS} 和 Ω_{CR}。源节点随机地从其集合中选择消息，然后在网络里进行传输：n 为分组码码长，h_{PS} 和 h_{CR} 为码率，f_{ij} 为链路 (i, j) 上的编码函数。在我们的 CRRN 中，中继节点对来自主系统和认知无线电的消息分别进行译码，然后将其叠加并加以传输。所以该链路用来中继认知无线电和主系统的编码函数分别为 $f_{ij,CR}$ 和 $f_{ij,PS}$，它们分别对认知无线电和主系统的消息进行编码。于是 $A_{ij}=A_{ij,CR}\times A_{ij,PS}$ 将在链路 (i, j) 上传输。根据以上设置，链路 (i, j) 上对认知无线电和主系统信息进行中继的码率为

$$R_{ij}=n^{-1}\log_2|A_{ij}|=n^{-1}\log_2|A_{ij,PS}|+n^{-1}\log_2|A_{ij,CR}| \tag{6-7}$$

这就是众所周知的叠加编码。这个编码方案将两套网络编码进行叠加。尽管由参考文献 [11] 可以推导出有时叠加编码并不能达到多源网络的容量上限，我们仍然采用该方案，因为对主系统和认知无线电进行联合压缩并编码的方案可能并不可行。如果主系统的网络容量保持不变，我们并不需要改变 CRRN 中的主系统网络编码。换言之，在 CRRN 中，主系统可使用原有的网络编码，在叠加编码下就好像认知无线电终端并接入网络一样。叠加编码是一种译码转发的协作策略。

6.1.1.3 在译码转发策略下的网络容量

在译码转发策略下，对于那些包含来自认知无线电和主系统的数据流的链路，我们可以将其链路容量写成 $R_{ij}=R_{ij,CR}+R_{ij,PS}$，其中 $R_{ij,CR}>0$，$R_{ij,PS}>0$。在我们的设置下，通过网络编码，可以通过计算最小割容量，分别对主系统和认知无线电推导出网络容量。我们用以下过程分析其干扰。首先，推导出原有网络

$G=(V, E)$，$\boldsymbol{R}=[R_{ij}]$ 的网络容量 v。然后，我们可以推出主系统的网络容量、认知无线电的网络容量，以及整个 CRRN 的容量 $v_t=v_{PS}+v_{CR}$。这样，我们就可以知道保证获得 $v_{PS}=v$ 和 $v_{CR}>0$ 的条件，这就意味着认知无线电的干扰是可以避免的。进一步地，我们想要通过适当地分配链路容量，在 $v_{PS}=v$ 的前提下最大化 v_{CR}。如果可行，我们则称之为干扰有界。

6.1.2 基本 CRRN 拓扑的网络容量分析

在不同的 CRRN 拓扑中，基于译码转发协作方式的认知无线电可以在不同程度上对主系统造成干扰。基于拓扑和链路容量，认知无线电的干扰是可避免的，也可能是不可避免的、可限的或者不可限的。我们在以下的基本拓扑下分析干扰的性质，并将其推广到任意的拓扑。我们把中继节点和源节点或者目的节点之间的链路的容量设为无限，这样我们就可以集中研究协作网络里的干扰了。

6.1.2.1 单跳中继网络

我们从最简单的拓扑——单跳中继网络开始分析。在这个网络里，只有一个链路。直观地看，既然认知无线电在传输，主系统将会被认知无线电干扰。如果认知无线电占据了整个链路的容量，主系统将无法进行传输。因此，在这个网络里，干扰是不可避免的，也是无法限制的。

6.1.2.2 汇接中继网络

一个汇接中继网络是由一系列串联的节点构成的。这样的网络的容量是由所有链路中容量最小的链路决定的。因此，认知无线电可以连接到网络中，利用那些容量更大的链路进行传输，即图 6-3b 中 b > a 的情况。除非我们给认知无线电分配了更多的容量，使得相应的链路成为主系统中容量最小的链路，认知无线电是不会对主系统发生干扰的，所以，干扰是可以避免的。但是，如果认知无线电占据了任何链路的全部容量，主系统将无法通过中继网络传输任何数据。因此，如果我们不对其加以限制，认知无线电的干扰是无界的。

6.1.2.3 协作式中继网络

类似于中继通信，我们通过加入协作式中继节点形成一个协作式中继网络（图 6-4 中的节点 2），于是网络中就有了三条链路以及两个割集。我们首先考虑原来的网络的容量（即在没有认知无线电的情况下图 6-4a 左边）。令链路（1，2）、（1，3）、（2，3）的容量分别为 a、b、c。主系统的割集为 {(1，2)(1，3)}、{(2，3}，(1，3)}。该主系统网络的网络容量为

$$v=\min(a+b,\ b+c) \tag{6-8}$$

现在我们计算协作式中继的网络容量，正如图 6-4a 右边所示。我们将认知无线电目的中继和节点 2（也就是协作式中继节点）连接。将认知无线电的网络容量记为 v_{CR}，主系统的容量记为 v_{PS}，CRRN 的总网络容量记为 v_t。认知无线电

网络的割集为（1，2）。主系统和认知无线电系统的总网络容量为

$$v_{PS}=\min(a-v_{CR}+b,\ b+c),\ v_{CR}\leqslant a \tag{6-9}$$

如果 $a+b>v$，v_{CR}可以在 $v_P=v$ 时大于零。在以下条件下

$$v_{PS}=\begin{cases}a-v_{CR}+b & v_{CR}\geqslant a-c\\ v & v_{CR}<a-c\end{cases},\ v_{PS},\ v_{CR}>0 \tag{6-10}$$

$$v_t=\begin{cases}a+b & v_{CR}\geqslant a-c\\ v+v_{CR} & v_{CR}<a-c\end{cases} \tag{6-11}$$

如果我们设置 $v_P=v$，我们知道

$$v_{CR}\leqslant a-c \tag{6-12}$$

因此 $a-c$ 是认知无线电在 $v_{PS}=v$ 的条件下的最大网络容量。注意到 $v_{CR}\leqslant a$，因此我们总有 $v_{PS}\geqslant b$。这就证明了在这种情况下，认知无线电网络对主系统的干扰是有界的。在这样的设置下，当我们把一个认知无线电源节点和一个认知无线电目的节点加到（2，3）中时，情形是一样的，只需在之前的分析里将 a 改为 c。

基于以上分析，如果认知无线电的消息是通过不属于割集的链路中继的话，认知无线电对主系统的干扰就是可以避免的。进一步地说，在这样的情况下，其干扰对主系统是有界的。这个现象来源于这样的一个事实：增加协作式中继节点，可以创造新的路由和更多的割集，因此认知无线电对主系统的干扰是可以避免的，而且是有界的。在这个情况下，认知无线电和主系统的网络容量间有两种可能性。当认知无线电的网络容量还没有达到其最大值时，我们可以增加其网络容量而不改变主系统的网络容量。在达到认知无线电的网络容量后，认知无线电和主系统的网络容量可以在一定范围内进行调节，可是我们赋予认知无线电的网络容量等于从主系统那里减下来的容量。反之亦然。

6.1.2.4 并行协作式中继网络

直到此刻，我们只考虑了一个认知无线电的目的节点。现在，我们把分析推广到多个认知无线电目的节点的情况来研究相应的干扰。在协作式中继网络里，我们加入一个额外的协作式中继节点，而在避免对主系统的干扰的前提下，只有一个认知无线电目的节点可以被加入。现在我们加入另外一个节点（图 6-4b 中的节点 3）来形成一个与原来的网络并行的协作式中继网络。接下来，我们在 CRRN 中加入另外一个认知无线电链路。在这样的拓扑里，我们有 4 个割集，每个割集有两条链路。

让我们再次考虑原主系统的网络容量，如图 6-4b 左边所示。将链路（1，2）、（1，3）、（2，4）、（3，4）、（1，4）的容量分别记为 a、b、c、d、e。主系统的割集为 {(1，2)，(1，3)，(1，4)}、{(2，4)，(3，4)，(1，4)}、{(1，2)，(3，4)，(1，4)}、{(1，3)，(2，4)，(1，4)}。该网络的网络容量为

$$v=e+\min(a+b,\ c+d,\ b+c,\ a+d) \tag{6-13}$$

我们接下来分析与认知无线电共存的并行协作式中继网络的容量。认知无线电网络的割集为 {1，2}、{1，3}，主系统和认知无线电的网络容量分别为

$$\begin{aligned} v_{PS} &= \min(a+b-2v_{CR},\ a-v_{CR}+d,\ b-v_{CR}+c,\ c+d)+e \\ v_{CR} &\leqslant \min(a,\ b) \end{aligned} \tag{6-14}$$

因此，如果 $a+b$、$a+d$、$b+c>v$，也就是，$a>c$、$b>d$，那么当 $v_{PS}=v$ 时 v_{CR} 可以大于 0。在此条件下，我们假设 $a+d>b+c$，可以得到

$$v_{PS}=e+\begin{cases} a+b-2v_{CR}, & v_{CR}>b-d \text{ 且 } v_{CR}>a-c \\ b-v_{CR}+c, & v_{CR}>b-d \text{ 且 } v_{CR}>a-c \\ c+d, & v_{CR}<b-d \text{ 且 } v_{CR}>a-c \end{cases} \tag{6-15}$$

$$v_{t}=e+\begin{cases} a+b-v_{CR}, & v_{CR}>b-d \text{ 且 } v_{CR}>a-c \\ b+c, & a-c>v_{CR}>b-d \\ c+d+v_{CR}, & v_{CR}<b-d \text{ 且 } v_{CR}<a-c \end{cases} \tag{6-16}$$

$a+d<b+c$ 的情况是对称的。如果我们设置 $v_{PS}=v$，有

$$v_{CR}=\min(a+d,\ b+c)-v \tag{6-17}$$

这个拓扑可以在两个并行中继节点中，容纳两个认知无线电的目的节点。认知无线电的干扰仍然是可以避免的，并且是可以限制的。我们可以推测出，由于协作式中继节点创造的并行结构，在这个并行协作式中继 CRRN 中，更多的认知无线电目的节点可以实现共存。可是，在两个认知无线电目的节点和一个主系统目的节点的情况下，增加认知无线电的网络容量可能会降低整个网络的容量。这时因为这两条与认知无线电目的节点连接的链路属于同一个割集，所以如果我们把认知无线电的网络容量增加 x，需要将主系统的容量减少 $2x$，这样整个网络容量就下降了 x。可是如果我们适当地选择认知无线电的容量的话，我们仍然可以提高总的网络容量，并且保证不对主系统产生干扰。

从以上分析我们可以看到干扰和网络拓扑之间的一些重要关系：

1）干扰避免：

在单跳中继网络中，干扰总是不可避免的，因为割集包括了所有的边。

在汇接中继网络中，如果认知无线电的割集没有包含在主系统的最小割集中，干扰是可以避免的。

2）有界干扰：

协作式中继网络中，干扰是有界的，因为认知无线电不可能占据任意主系统割集的所有的边。

3）多个认知无线电目的节点：

在有多个认知无线电目的节点的并行协作式中继网络中，以上性质仍然成立。可是当放宽避免干扰的限制时，CRRN 中总的网络容量可能降低。

现在我们将以上现象推广到任意的译码转发网络拓扑中，在以下分析中，我们将认知无线电的网络容量记为 v_{CR}，主系统的总网络容量记为 v_{PS}，原始的网络容量记为 v，CRRN 的总网络容量为 $v_t = v_{PS} + v_{CR}$。我们首先阐述可避免的干扰，然后放宽避免干扰的限制，分析不可避免的干扰，最后推导出有界干扰的条件。

关于可避免的干扰，在绝大多数情况下，认知无线电试图在避免对主系统产生干扰的情况下使用通信资源。因此，我们强调这一方面，并在以下提供关于认知无线电如何在利用主系统进行中继时避免干扰的细节。引理 6.1 推导出了针对中继网络拓扑的条件，基于引理 6.1，定理 6.1 进一步推导出针对 CRRN 的条件。在这些网络拓扑中，我们在定理 6.2 里推导出单播主系统中认知无线电的最大网络容量，并且提供在单播和多播主系统中达到最大容量的算法。最后，我们简短地分析当认知无线电开始干扰主系统时网络容量的变化。

引理 6.1：假定每一条链路都属于 CRRN 中至少一个割集，如果 $v_{PS} = v$，那么 $v_{CR} = 0$。

证明：记主系统的割集集合为 φ，φ 包括所有的主系统的割集，其容量等于 v。一个割集 c 在原来的网络里的容量记为 $R(c)$，在 CRRN 里的容量记为$R'(c)$。假设 $U_{c_i \in \varphi} C_i = E$。如果 $v_{CR} > 0$，那么至少有一条链路使得 $R_{ij,CR} > 0$。在边（i，j）上的主系统链路容量变为

$$R'_{ij,PS} = R'_{ij} - R_{ij,CR} = R_{ij} - R_{ij,CR} < R_{ij} \tag{6-18}$$

这个边（i，j）必须属于某个割集 $C_{i \subset \varphi}$。因为 $R'_{ij,PS} < R_{ij}$，我们有 $R'(c_i) < R(c_i)$。因此，我们可推断 $v_{PS} = R'(c_i) < R(c_i) = v$。由此我们得出矛盾。引理得证。

定理 6.1：假设 $v_P = v$，那么 $v_{CR} > 0$，当且仅当在包括所有主系统最小割集的边集合里没有割集。

证明：我们首先证明“当”部分。令链路集合 $\varphi_P = \{e_i | e_i$属于任何容量等于 v_{PS}的割集$\}$。令割集集合 ψ_{CR}包括认知无线电中的所有割集，令 h_{PS}为主系统割集中不属于 φ_{PS}的最小主系统割集容量。如果对于所有 $c_i \in \psi_{CR}$，$c_i \not\subset \varphi_{PS}$，给认知无线电的每个满足 $e_i \notin \varphi_P$ 链路分配最小容量 $h_{PS} - v_{PS}/m > 0$。令 m 为中继网络里的边的数量。在同一割集中边的数量不会超过 m。所有 ψ_{CR}中的每个元素的容量都大于或等于 $h_{PS} - mh_{PS} - v_{PS}/m$。这里

$$v_{PS} = v，\ v_{CR} \geqslant h_{PS} - v_{PS}/m > 0$$

现在我们证明“仅当”部分。假定 $v_{CR} = \Delta > 0$，那么 ψ_{CR}中的每个元素的容量大于或者等于 Δ。可是 $v_{PS} = v$，所以 φ_P中的每个元素的链路容量无法给认知无线电分配容量，所以对于所有 $c_i \in \psi_{CR}$，$c_i \not\subset \varphi_{PS}$。

注意到当我们考虑压缩中继的协作策略，而非译码转发时，“仅当”部分的结论未必成立。

接下来我们推导在避免对主系统干扰的限制下，认知无线电的最大网络容量，以及在主系统单播网络中的链路容量分配方法。为了简化问题，在我们推导认知无线电最大网络容量时，首先只考虑一个主系统目的节点和一个认知无线电目的节点。我们定义 CRRN 的割集为将 CRRN 分为两个不相连的网络的边集合：一个包含着认知无线电源节点和主系统源节点，一个包含认知无线电和主系统的目的节点。于是，CRRN 的割集必然包括认知无线电的割集以及主系统的割集。我们假设中继网络中的边包含在至少一条从主系统源节点到主系统目的节点的路径里。根据这一假设，每个 CRRN 的割集都是主系统的割集。

定理 6.2：令 $v_{PS}=v$。CRRN 中认知无线电的最大网络容量为 CRRN 的最小割集容量减去原始网络的容量。

证明：令 CRRN 的最小割集容量为 v_{CRRN}。我们可以加上一个连接认知无线电源节点和主系统源节点的超级源节点，以及一个连接认知无线电目的节点和主系统目的节点的超级目的节点，这样，超级源节点的网络容量为 v_{CRRN}，并且 $v_{PS}+v_{CR}=v_{CRRN}$，因此

$$v_{CR}=v_{CRRN}-v_{PS}=v_{CRRN}-v \tag{6-19}$$

定理得证。

6.1.3 链路分配

在推导出认知无线电的最大网络容量后，我们应该找到达到认知无线电最大网络容量的网络容量分配方法，而且保证不对主系统产生干扰。CRRN 中的链路分配问题可以表达为一个多商品流问题。来自认知无线电和主系统的信息即商品，认知无线电网络容量 $v_{CRRN}-v$ 和主系统的网络容量 v 为商品需求。所以，我们可以利用线性规划，解决这一多商品流问题，实现对链路容量的分配，以达到认知无线电的最大网络容量。

我们放弃单播的假设，令主系统为多播，而认知无线电仍然为单播。对于一个多播网络，网络容量是由其最小割集容量限制的。所以我们可以通过推广多商品流问题表达认知无线电最大网络容量的问题。

命题 6.1：在包括多播主系统和单播认知无线电的 CRRN 中，达到最大网络容量的链路分配问题可以表达成一个线性规划问题。

以下我们证明命题 6.1 的正确性。我们分别考虑从主系统源节点到每一个目的节点的数据流。每一个流应该超过或者至少等于 v，以确保 $v=v_{PS}$。于是我们在每一条边上选择在这些数据流中最大的流量，将其分配给主系统，并最大化认知无线电的网络容量。

现在我们可以证明这个问题也可以表达成一个线性规划问题。我们将原网络容量的符号从 v 改成 α，以区分顶点 v 的符号。于是在主系统多播网络中，我们

的链路分配算法可以表达如下：S_{PS}和S_{CR}代表主系统和认知无线电的源，t_{PSl}和t_{CR}代表主系统和认知无线电的目的节点（主系统有多个目的节点），f_{PSi}和f_{CR}代表分配给主系统和认知无线电的链路容量，v代表中继节点。在主系统多播的CRRN中，链路分配的线性规划问题如下：

最大化：

$$\sum_{v \in V_{CRs}} f_{CR}(s,v), V_{CRs} = V - \{t_{PSl} \mid l = 1,\cdots,k\} \tag{6-20}$$

满足：

(i)
$$\sum_{v \in V_{is}} f_{PSi}(S_{PS},v) \geqslant a$$

$$V_{is} = V - \{S_{CR},\ t_{CR},\ t_{PSl} \mid l = 1,\ \cdots,\ k\} \tag{6-21}$$

(ii)
$$\sum_{v \in V_i} f_{PSi}(u,v) = 0$$

式中，
$$V_i = V - \{t_{CR},\ t_{PSl} \mid l = 1,\ \cdots,\ k,\ l \neq i\}, \tag{6-22}$$

$$u \in V - \{S_{CR},\ s_P,\ t_{CR},\ t_{PSl} \mid l = 1,\ \cdots,\ k\}$$

(iii)
$$\sum_{v \in V_{CR}} f_{CR}(u,v) = 0$$

式中，
$$V_{CR} = V - \{t_{PSl} \mid l = 1,\ \cdots,\ k\}, \tag{6-23}$$

$$u \in V - \{S_{CR},\ s_{PS},\ t_{CR},\ t_{PSl} \mid l = 1,\ \cdots,\ k\}$$

(iv)
$$f_{PSi}(u,\ v) = -f_{PSi}(u,\ v) \tag{6-24}$$

(v)
$$f_{CR}(u,\ v) = -f_{CR}(u,\ v) \tag{6-25}$$

(vi)
$$f_{PSi}(u,\ v) \leqslant c(u,\ v) - f_{CR}(u,\ v),\ i = 1,\ \cdots,\ k \tag{6-26}$$

我们可以用多种众所周知的方法来解决以上的线性规划问题，例如单纯形方法。

以上的结果提供了一个构成CRRN的方法，其中认知无线电可以避免对主系统的干扰。引理6.1证明了原始网络不应该在网络的瓶颈上有链路。最小割集即网络的瓶颈。定理6.1推导出了认知无线电应该满足的条件：主系统的最小割集不应该包括认知无线电的任何割集。也即，我们应该可以找到认知无线电的至少一条从源到目的的路由，不经过主系统的瓶颈。引理6.1和定理6.1给我们提供了CRRN拓扑的全面特性。给定CRRN的拓扑，并且假设主系统和认知无线电都是单播，定理6.2给出了在避免对主系统干扰的限制下，认知无线电的最大网络容量，这是认知无线电可以通过中继网络获得的，以网络容量衡量的资源的上界。接下来，我们把达到认知无线电最大网络容量的链路分配问题表达为多商品流问题，并可以用线性规划解决。最后，把我们的工作推广到多播主系统和单播认知无线电的情况，证明了认知无线电的最大网络容量问题仍然是一个线性规划问题，因此可以用相应的方法解决。

我们现在研究在认知无线电对主系统造成干扰时，认知无线电网络容量变化时随之而来的总网络容量的变化。根据定理6.2，认知无线电的最小割集必须包含在主系统的最小割集内，这也是CRRN在获得最大网络容量时的最小割集。因此链路容量分配的微小变化，在不影响认知无线电和主系统的割集的位置时，不会改变整个网络的容量。总网络容量仍然是CRRN的最小割集容量。可是当我们考虑认知无线电是多播时，链路容量分配的变化可能会减小总网络容量，正如以下引理所言。

引理6.2：如果CRRN中，认知无线电是多播，而主系统是单播，增加认知无线电的网络容量会减小网络的总容量，如果不同的认知无线电目的节点的不同割集占据了不止一条属于同一主系统最小割集的链路。

证明：我们只证明两个认知无线电目的节点并占据属于同一主系统最小割集的两条链路的情况，因为其他情况都只是这个证明的简单推广。将这两条属于不同认知无线电的最小割集的链路记为e_1、e_2，这两条链路属于同一个主系统割集φ_1。令链路e_i的容量分配为$c_{ei}=c_{ei,\mathrm{PS}}+c_{ei,\mathrm{CR}}$。现在，如果我们增加在$e_1$和$e_2$上分配给认知无线电的网络容量，增量为$\delta>0$，有

$$C_{eI}=(c_{eI,\mathrm{PS}}-\delta)+(C_{eI,\mathrm{CR}}+\delta),\ I=1,\ 2 \tag{6-27}$$

因为e_1和e_2在认知无线电的不同目的节点的最小割集上，属于主系统的同一个最小割集，有

$$v'_{\mathrm{CR}}\leqslant v_{\mathrm{CR}}+\delta \tag{6-28}$$

$$v'_{\mathrm{PS}}=v_{\mathrm{PS}}-2\delta \tag{6-29}$$

$$v'_t=v'_{\mathrm{CR}}+v'_{\mathrm{PS}}\leqslant v_{\mathrm{CR}}+v_{\mathrm{PS}}-\delta<v_t \tag{6-30}$$

式（6-28）中的不等式是因为存在认知无线电另外的割集，不包括e_1和e_2。根据式（6-30），CRRN的总的网络容量至少减少了δ。

接下来我们研究当干扰不可避免时，认知无线电对主系统的干扰。我们把讨论限制在只给认知无线电分配较小容量的情况，估计对主系统干扰的上界。以下引理的想法以及证明都来源于定理6.1的证明，仅做了一些修改。

引理6.3：假设认知无线电的干扰是不可避免的。令链路集合$\varphi_{\mathrm{PS}}=\{e_i|e_i$属于任何容量等于$v_{\mathrm{PS}}$的割集$\}$。认知无线电中有$m$个割集包含于$\varphi_{\mathrm{PS}}$。那么在$\varphi_{\mathrm{PS}}$不受分配给认知无线电的链路容量的限制的情况下，认知无线电对主系统的网络容量干扰的上界为$V_{\mathrm{CR}}m$。

证明：假设认知无线电在φ_{PS}中的割集不交叠。我们将链路容量分配给认知无线电，获得较小的网络容量v_{CR}，并且不改变φ_{PS}。因此，主系统的最小割集仍然包含在原来的φ_{PS}中，φ_{P}中有m个认知无线电的割集获得了链路容量v_{CR}。假设这些认知无线电的割集都只包含在一个主系统的最小割集中，那么这个割集的容量将减少$v_{\mathrm{CR}}m$，主系统的网络容量也成比例地减少。如果我们放弃认知无线

电割集不交叠的假设，或者干扰只在主系统的最小割上，那么干扰将被分散，使得干扰对网络容量的影响减小。因此，在我们的假设下，干扰的上界将是 $v_{CR}m$。

现在我们研究有界干扰的条件。

引理 6.4：如果所有认知无线电路径的边集合不包括在任何主系统的割集中，认知无线电对主系统的干扰是有界的。

证明：我们称占据了所有从源节点到目的节点路径的链路的认知无线电将给主系统带来最大危害。如果所有认知无线电的路径的边集合不包括任何主系统的割集的话，即便是在最坏的情况下，主系统的最小割集容量将大于 0。因此，所有主系统的网络容量总是大于 0，换言之，认知无线电的干扰是有界的。

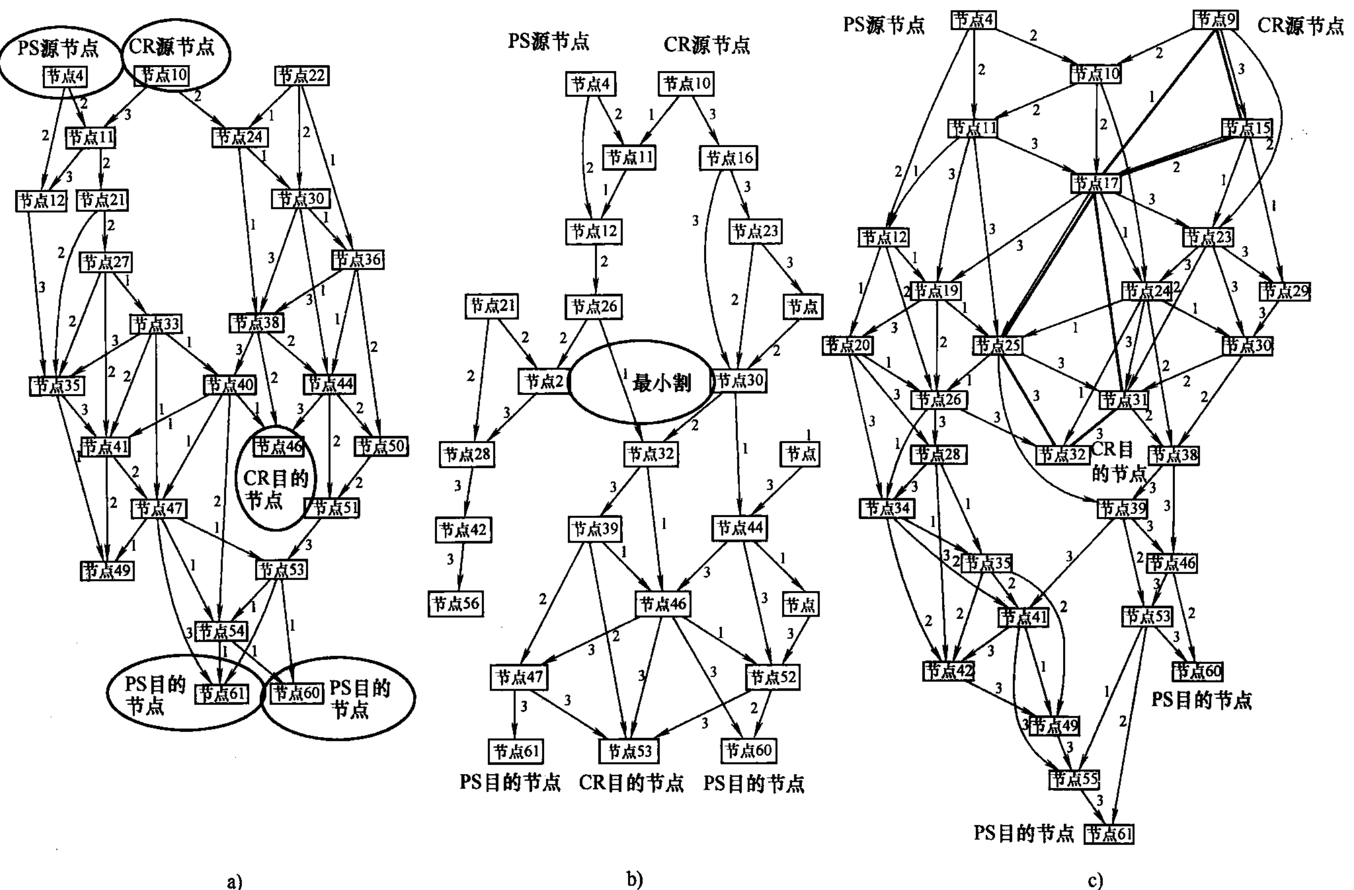
在满足以上条件的网络中，主系统总是能够保持传输，即便认知无线电违反访问规范。

6.1.4 数值结果

为了验证我们的算法可以应用到随机生成的一般 CRRN 拓扑中，我们在图 6-5a 所示的 7 × 7 网格中，对一个随机生成的 CRRN 进行了仿真。中继节点以概率为 0.5 出现在图的每一个格点上。我们考虑单源节点和双目的节点的多播主系统，以及单源节点和单目的节点的单播认知无线电，如图 6-5a 所示。

仿真参数设置如下。每条链路的容量从区间［1，3］中随机选取一个整数。如果两个节点在网格中的距离小于 2，我们就认为这两个节点间存在一条链路。我们的算法对随机生成的 CRRN 拓扑运行 1000 次。认知无线电的最大网络容量占主系统的网络容量的比例直方图如图 6-6 所示。我们用比例来比较主系统的通信资源和认知无线电的通信资源。当认知无线电的最大网络容量为 0 时，比例定义为 -0.5。仿真结果表明，在随机生成的 CRRN 拓扑中，认知无线电可以利用主系统来传递数据包的概率为 92%。我们也注意到，认知无线电容量与主系统网络容量相同或者略少的情况经常发生。从平均意义上讲，认知无线电的网络容量为主系统容量的 1.3 倍。通过这个结果，我们可以大致知道，当认知无线电利用主系统网络进行中继时，其最大网络容量的概率分布。

认知无线电获得高网络容量时，通过研究 CRRN 的拓扑，我们观察到主系统的网络容量被一个或者几个小的割集所限制，如图 6-5b 所示，或者认知无线电的源节点和目的节点的距离较近，如图 6-5c 所示（如图中所示，认知无线电只需要在中继网络中传输一跳或者两跳，就可以到达其目的节点）。我们可以直观地从理论结果中推测出这些特性。如果主系统的最小割集的数量较小，其容量也较小，主系统就只能使用网络里的一小部分资源。于是，认知无线电就有了更多地利用这些资源的机会。另一方面，如果认知无线电只使用 CRRN 中较少的几跳，它就可以利用这些跳更多的资源，如果这些跳不是主系统的瓶颈。在这些



a)　　b)　　c)

图 6-5　仿真场景

a）随机产生的 CRRN 拓扑　b）主系统的网络容量被一个小的最小割限制

c）CR 之间的小距离（这里，源节点和目的节点间中继网络里两跳中的一个）

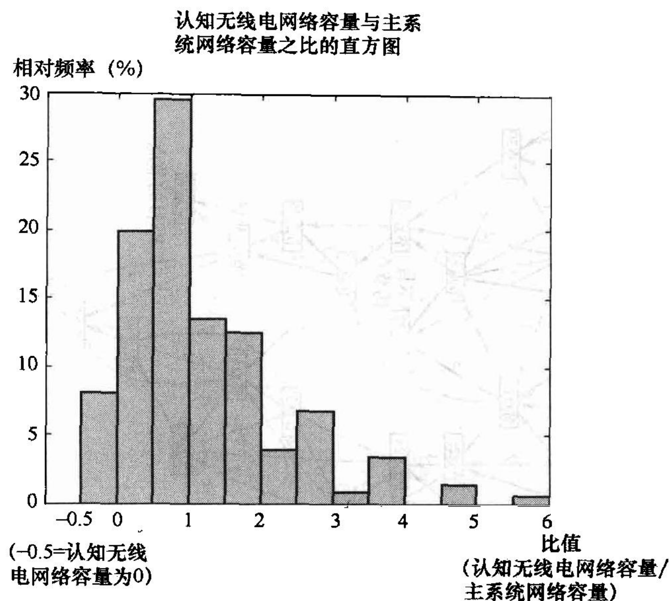


图 6-6 容量增益直方图

情况里，我们可以分配更多的容量给认知无线电，而无需对主系统造成干扰。因为定理 6.1 的条件，或者认知无线电的源节点和目的节点之间没有路由，认知无线电的最大网络容量有可能为 0。

我们本节分析了当 CRRN 中认知无线电使用主系统进行中继时，对应于网络拓扑的干扰的特性。当认知无线电使用主系统进行中继时，我们改动译码转发协作策略。我们从基本的网络拓扑开始，研究了认知无线电的干扰是否可以避免或者是否有界。我们推导出了如下结论：干扰在单跳中继里是可以避免的，在汇接网络里是无界的，而在协作式中继结构里是有界的。然后，我们把工作推广到多认知无线电目的节点的情况。在推广理论结果时，因为认知无线电对主系统干扰避免的本质，我们着重强调了对干扰避免的分析。我们首先推导出了推广的 CRRN 的条件，以保证认知无线电的干扰是可避免的：主系统的最小割集不能包括认知无线电的任意割。因此，在设计 CRRN 时，我们应该至少有一条认知无线电的路径绕开主系统的瓶颈。在这种拓扑里，我们推导出在主系统单播的 CRRN 中，保证避免对主系统的干扰的限制下，认知无线电的最大网络容量。我们也证明了为获得最大网络容量的链路容量分配，在单播和多播的主系统中，可以表达成一个线性规划问题。当认知无线电超越干扰的限制时，我们描述了对应于认知无线电网络容量的总网络容量的变化。我们接下来分析了不可避免的干扰，并估计了在认知无线电网络容量较小时干扰的界限。最后，我们推出了认知

无线电干扰有界的 CRRN 拓扑的推广条件。我们对随机产生的 CRRN 拓扑进行了仿真，运行了链路分配算法，来获得认知无线电的最大网络容量。对平均意义而言，认知无线电的最大网络容量是主系统的 1.3 倍，这是一个很大的增益。通过在 CRRN 中分析干扰的特性，路由和调度可以建立起来以提高网络应用 CRRN 的效率。因此，使用主系统进行数据包中继的认知无线电系统被证明可以在网络吞吐量和高可利用度上获得频谱效率，认知无线电网络的基础也就随之可建了。

6.2 认知无线电网络体系结构

认知无线电可以在链路层提高频谱利用率这一事实已经广为人知了。我们也证明了认知无线电和主系统节点间的协作式中继，可以通过构造一个广义的认知无线电网络以极大地增加网络容量，这意味着认知无线电将感知其附近可用的网络和通信系统来完成网络功能，而不仅限于在链路层利用频谱空洞。因此，认知无线电网络不只是另外一种以认知无线电连接的网络。它们包括了不同种类的共存的多无线电通信系统，其中包括认知无线电系统。认知无线电网络可以被看成某种包括多种通信系统的异构网络。其异构性来自于各种无线接入技术、网络、用户终端、应用，以及服务运营商等。认知无线电的体系结构设计的目标是提高网络的利用率。从用户的角度看，网络的利用率就意味着它们可以在任何时间和任何地点通过接入认知无线电网络来满足其需求。从运营商的角度看，它们不仅可以为移动用户提高更好的服务，还可以更有效地分配无线电和网络的资源。

6.2.1 网络体系结构

认知无线电可以使用在中心网络、分布式网络、Ad hoc 和网状结构中，满足有执照和无执照的应用需求。认知无线电网络的基本构件是移动终端（MS）、基站/接入点（BS/AP）和骨干网/核心网。这三个基本构件组成了认知无线电网络里三种网络体系结构，即基础体系结构、Ad hoc 体系结构和网状体系结构。

6.2.1.1 基础体系结构

在基础体系结构中（见图 6-7），一个移动终端只能以单跳的方式接入一个基站或者接入点。在同一个基站或者接入点的传输范围内，移动终端应该通过基站或者接入点进行通信。在不同小区间进行的通信应该通过骨干网/核心网进行。基站/接入点可以运行一个或者多个通信标准/协议来满足移动终端的不同需求。一个认知无线电终端也可以通过它们的基站/接入点接入不同的通信系统。

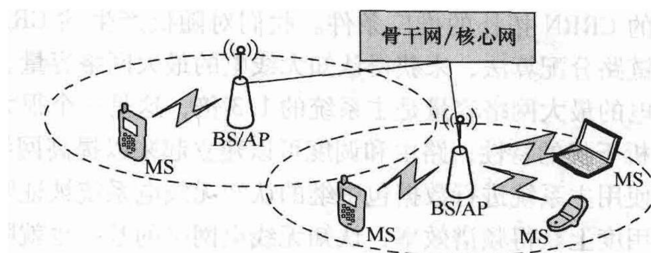


图6-7 认知无线电网络中的基础体系架构

6.2.1.2 Ad hoc 体系结构

在 Ad hoc 体系结构内，没有基础体系结构的支持（见图6-8）。这种网络是实时动态生成的。如果一个移动终端发现附近有其他移动终端，而且可以通过一定的通信标准/协议进行连接，它们就可以设置一条链路，形成一个 Ad hoc 网络。注意，这些节点间的链路可能是由不同通信技术完成的，并且两个认知无线电终端，可以使用现存的通信协议（例如 WiFi、蓝牙），或者动态地使用频谱空洞来进行通信。

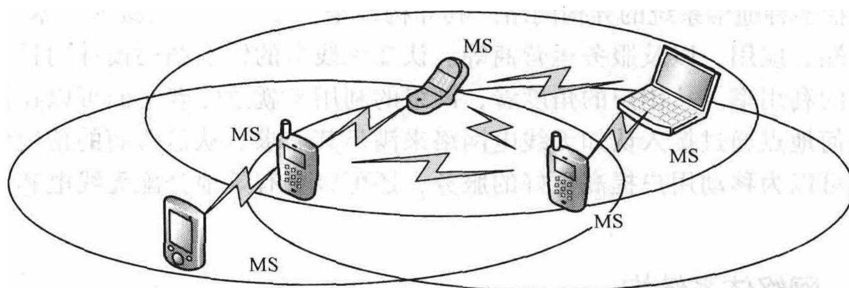


图6-8 认知无线电网络中的 Ad hoc 体系结构

6.2.1.3 网状体系结构

这种体系结构是基础体系结构和 Ad hoc 体系结构的结合，并且保证基站/接入点之间的无线连接（见图6-9）。这种网络体系类似于混合无线网状网。在这种体系结构里，基站/接入点如同是无线路由器，形成无线骨干网。移动终端可以直接接入基站/接入点，或者利用其他移动终端作为多跳中继节点。有些基站/接入点可以连接到有线骨干网/核心网络并作为网关。既然基站/接入点无需连接到有线骨干网/核心网，这就增加了适应性，并减少了对基站/接入点的位置规划的费用。如果基站/接入点有认知无线电的功能，它们可以通过频谱空洞互相进行通信。因为现在频谱利用率的低下，可能存在大量频谱空洞可以利用。因此，在认知无线电基站/接入点之间的无线通信链路的容量可能会很大，这使得无线

骨干网络可以提供更多的数据流量。

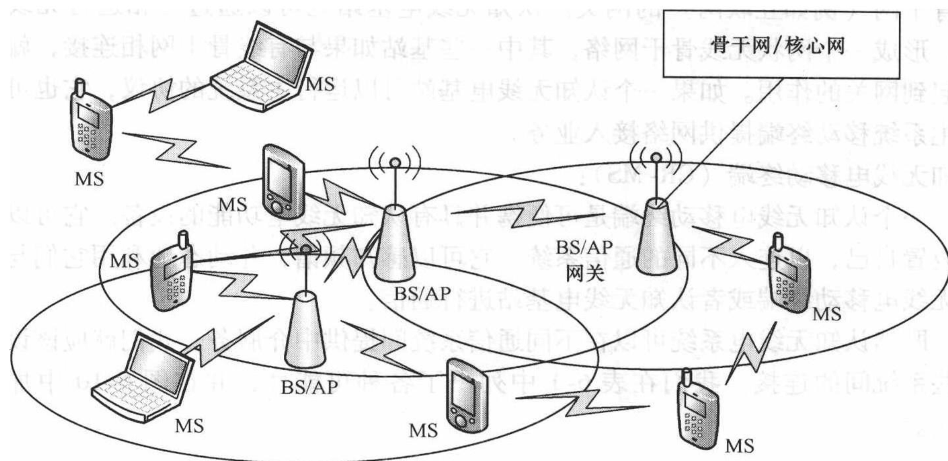


图6-9 认知无线电网络中的网状体系结构

6.2.2 认知无线电网络中的链路

我们可以回顾一下认知无线电网络中的两种无线通信系统：主系统（PS）和认知无线电（CR），其区别为频段上的不同优先级。一个主系统是指现存的，并在一个或者多个固定频段上工作的系统。不同的主系统在有执照或者无执照的频段上工作，或者在相同的地理位置，或者在相同的频段（或者是相同的频段集合）。它们可以描述如下：

工作在有执照频段的主系统：在有执照频段工作的主系统有使用该频段的最高优先级（例如2G/3G蜂窝系统、数字电视广播）。其他无执照用户/系统既不能无法容忍地干扰主系统，也不能占用有执照频段。

工作在无执照频段的主系统：在无执照频段上工作的主系统被称为无执照频段主系统。不同的主系统可以利用频谱兼容。特别地，在同一无执照频段上工作的主系统应该考虑相互的影响，实现共存。这些系统也许有不同的优先级，这取决于一些相关规定。

一个认知无线电系统无权接入某一固定频段。认知无线电系统必须通过利用频谱空洞和动态接入进行相互的通信。认知无线电系统由两部分组成：认知无线电基站（CR-BS）和认知无线电移动终端（CR-MS）。

认知无线电基站（CR-BS）：

一个认知无线电基站是认知无线电系统的固定组成，有认知无线电的功能。它代表认知无线电系统的基础体系结构的部分，并为认知无线电移动终端提供支

持（例如频谱空洞管理、移动管理、安全管理等）。它为认知无线电终端提供接入骨干网（例如互联网）的网关。认知无线电基站也可以通过互相进行无线通信，形成一个网状无线骨干网络，其中一些基站如果与有线骨干网相连接，就可以起到网关的作用。如果一个认知无线电基站可以运行主系统的协议，它也可以为主系统移动终端提供网络接入业务。

认知无线电移动终端（CR-MS）：

一个认知无线电移动终端是可便携并具有认知无线电功能的设备。它可以重新设置自己，以连入不同的通信系统。它可以感知频谱，并动态地利用它们与认知无线电移动终端或者认知无线电基站进行通信。

既然认知无线电系统可以在不同通信系统间提供中介服务，我们就应该设置一些系统间的连接。我们在表 6-1 中列出了各种可能性，并在图 6-10 中加以描述。

表 6-1　认知无线电网络中的链路总结

Rx/Tx	CR-MS	CR-BS	PS-MS	PS-BS
CR-MS	·	·	·	·
CR-BS	·	·	·	
PS-MS	·	·	·	·
PS-BS	·		·	·

· 表示可能存在链路

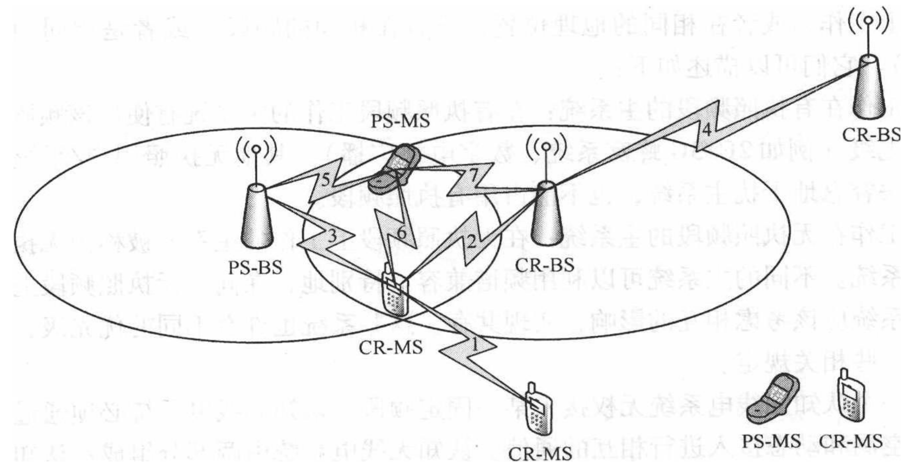


图 6-10　认知无线电网络中的链路

认知无线电移动终端↔认知无线电移动终端：

一个认知无线电移动终端可以与其他认知无线电移动终端进行直接通信。它

们可以协作地对不同有执照和无执照频段中的频谱空洞进行感知，并利用它们作为工作频段。需要一个共同的控制信道来保证它们可以互相交换信息。

认知无线电移动终端↔认知无线电基站：

一个认知无线电基站可以动态地感知附近一个可用的频段，并收集其他移动终端的感知结果，在其覆盖区域为认知无线电移动终端提供单跳接入。这或许需要协作式感知的技术。在认知无线电基站的协作下，认知无线电移动终端可以接入骨干网，或者与其他通信系统进行通信。

认知无线电移动终端↔主系统基站：

如果有需要让认知无线电移动终端和主系统基站相连接，认知无线电移动终端可以对自己进行重新配置，成为主系统的一部分（即主系统移动终端）。在这个情况下，它将在该频段上成为一个主用户。

认知无线电基站↔认知无线电基站：

当在认知无线电基站间实现直接无线连接时，它们可以形成一个网状无线骨干网。因为其认知无线电的能力，它们可以动态地选择工作频段并互相通信。既然认知无线电基站有更多的无线接口，认知无线电基站间的链路容量可能会很大。这种链路的另外一个好处就是减小开销。这是因为我们可以在一些有线连接不能实现的情况下部署认知无线电基站。

主系统移动终端↔主系统基站：

这是在移动终端和基站间典型的单跳连接。主系统基站负责在其覆盖区域内协调通信，并为主系统移动终端提供骨干网接入。这样的链路随时都是双向的，这是与其他种类的链路本质的区别。

主系统移动终端↔认知无线电移动终端：

为了提供不同通信系统间协同工作的能力，这种链路是必需的。在这种情况下，认知无线电终端应该重新设置自己，成为主系统的一部分。

主系统移动终端↔认知无线电基站：

为了提供不同通信系统间协同工作的能力，这种链路是必需的。如果认知无线电基站可以运行主系统的协议，它就可以为主系统移动终端提供接入服务。

主系统移动终端↔主系统移动终端：

这种通信在无线联网系统中可以 Ad hoc 网络的形式存在于主系统中。但是，在某些系统的基础体系结构模式下，这样的通信可能是被禁止的。不过，如果两个节点都可以转换成认知无线电，这就转换成了认知无线电移动终端之间的通信。

请注意以上列出的认知无线电链路的一个特点。除了主系统移动终端和主系统基站间的连接可以保证双向性外，其他七种连接在一次频谱接入内只能保持单向性。这不难理解，既然频谱机会在时间上可能很短，无法保证双向数据包的交

流，下一次的频谱机会也无法保证。这样的单向链路性质在考虑各种网络操作时，例如网络安全性，将起到关键的作用。这一点将在第9章中讨论。

6.2.3 认知无线电网络中的IP移动性管理

正如我们所提到的，认知无线电网络在很多方面是一种异构网络，例如无线通信技术和协议、骨干网络种类、用户终端种类、网络运营商等。一个认知无线电移动终端选择最好的通信系统来满足其需求。既然不同的无线系统有不同的媒体接入控制（MAC）和物理层（PHY），如何综合这些系统来为上层提供更好的服务，就成为了一个重要的任务。既然网络层是点到点的通信接口（或者接入技术）和端到端的层（例如传输层和应用层）之间的接口，它在综合的过程中起着至关重要的作用。

互联网协议（IP）及其推广已经被看作是将各种异构网络融合成单一的、基于全IP的、综合的网络平台，移动IP被看作更为相关的推广。在移动IP中有两种重要的实体：本地代理（HA）和外地代理（FA），它们分别是本地或者外地网络上的互联网路由器。一个移动节点（MN）通过HA或者FA接入互联网。与MN连接的节点被称为通信节点（CN）。我们希望移动IP能够在认知无线电网络里保持其在移动Ad hoc网络中起到的重要作用。

在传统的基础体系结构里，在基站和移动终端间只有一个跳跃，移动终端之间没有连接，也不允许多跳路径。既然所有的移动终端直接和基站连接，一个集中式的移动管理方案便成为可行的。例如在移动IP网络中，所有的基站可以通过代理广告，直接向移动终端广播它们的转交地址（CoA）。一个移动终端很容易从外地网络获得CoA，并向本地代理（HA）注册。然后，HA可以以隧道方式将所有数据包发给移动终端的最新的CoA。可是，在网状网络结构的认知无线电网络里，移动管理是更加具有挑战性的研究任务，特别是移动终端间有多跳中继功能时。相关的问题包括位置管理和切换管理。

6.2.3.1 位置管理

位置管理是一个两阶段的过程，使得网络可以找到移动终端的附着点以传递呼叫。这两个阶段包括位置注册和呼叫传递。当一个移动终端访问一个外地网络（FN），并希望得到互联网接入服务时，它首先通过检测代理广告找到FN的移动代理。在得到代理广告后，移动终端便可形成一个CoA并通知HA关于当前CoA和MN之间的联系。可是，在认知无线电网络中，一个认知无线电移动终端可以同时连接到多个不同的无线系统，这些系统有可能属于不同的FN，它应该从每一个系统获取一个CoA，以通过它们获取或者传输数据。所以，有必要开发新的策略来处理多重CoA。特别地，一个认知无线电移动终端可以从每一个连接的FN处获取多个CoA，以便认知无线电终端无需再用一个CoA来代表其当前的

位置以及传递数据包。此外，多个认知无线电终端可以形成一个 Ad hoc 网络，其中有些终端可以连接到基站/接入点以接入骨干网/核心网。我们把这些节点称为网关节点。因为基站/接入点的有限覆盖，有些移动终端只能通过多跳中继来得到基站的服务，如图 6-11 所示。这样的移动终端和 Ad hoc 网络的协作与融合是一项具有挑战性的任务。

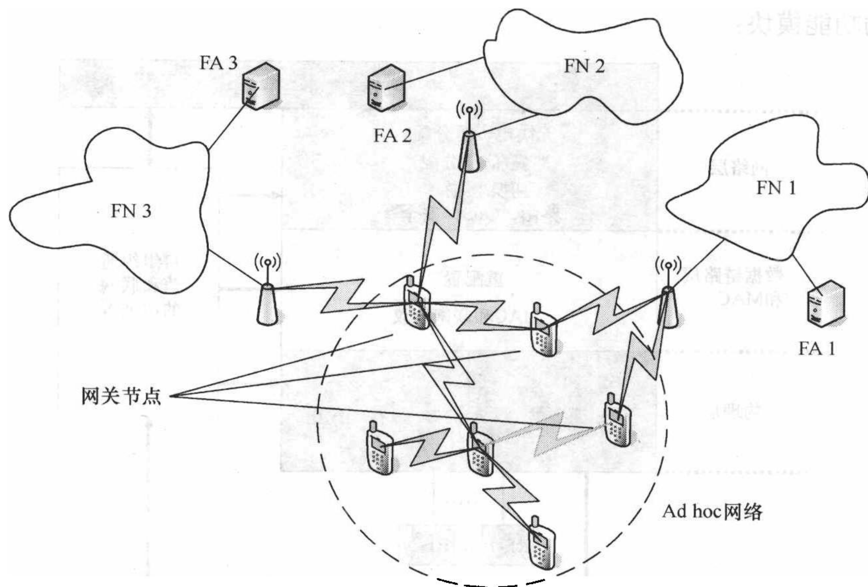


图 6-11 Ad hoc 网络和 IP 移动管理的结合

6.2.3.2 切换管理

切换管理使得网络可以保持用户在移动并改变网络接入点时的连接。这包括三个阶段：初始化、新连接的产生、数据流控制。因为认知无线电网络的多跳特性，切换管理不再是单个移动终端和 FN 之间的问题。它涉及到多个移动终端和 FN。例如，如果有的网关节点从基站的覆盖区域移走，它们应该通知 Ad hoc 网络中的节点连接已经失去。所以 Ad hoc 网络中的节点可以准备实施切换，如果它们和这些网关节点有连接的话。此外，由于网关节点的丢失，某些 FN 不再可连接。这能产生使 FN 无效的 CoA。新的 CoA 注册机制可能还是需要的。

6.3 认知无线电网络的终端体系结构

认知无线电及网络的实现是基于智能终端设备的，这些终端设备具有强大的处理分布式联网功能的能力。与传统的蜂窝通信及网络结构相比，认知无线电网

络包括了大量局域的自组织的最优化，从而在链路层和网络层达到全局频谱效率[21]。

6.3.1 认知无线电设备体系结构

图6-12描述了我们提出的自组织认知无线电的设备体系结构，它包括以下主要的功能模块：

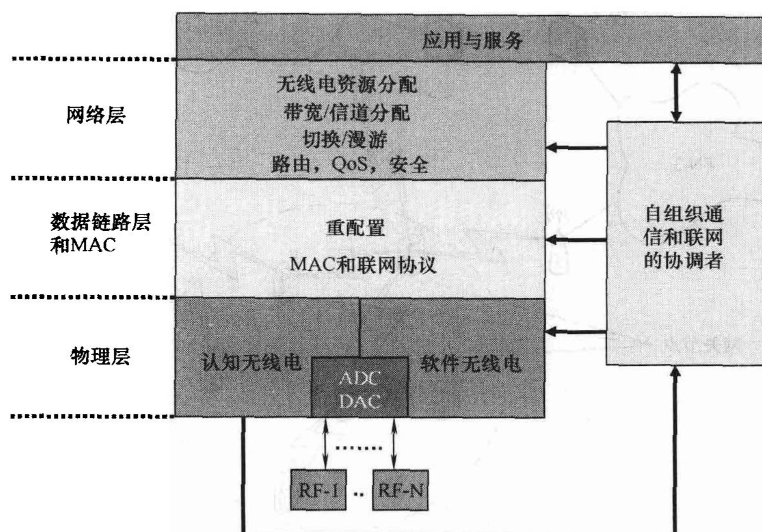


图6-12 认知无线电的终端体系结构

认知无线电：认知无线电识别无线通信环境以及共存的系统和网络。

软件无线电（SDR）：基于自协调者的决定，软件无线电为移动设备通信设置适当的接收机的参数。第2章提供了一些全面可编程的软件无线电的例子。

可重置MAC：自协调者也确定在可用的系统和网络间最优的路由，可重置MAC在一个全面可接入的协议机上调节合适的子程序。

网络层过程：自协调者也指导网络层的功能，如无线资源分配、移动管理等，来实现无线网络的运行。

自组织通信/联网协调者：终端设备的“大脑”决定：1）基于认知无线电信息的接入网络路由；2）适当的认知无线电硬件和软件的配置；3）用户对终端设备通信需求的维护。

射频：射频部分可能包括多个子带射频，以覆盖正确的频段，并且有可调节的射频滤波的能力，以满足选定的系统参数。

认知无线电和自组织协调者以及软件无线电之间的关系如图6-13所示。我

们暂且不考虑电路的复用情况。

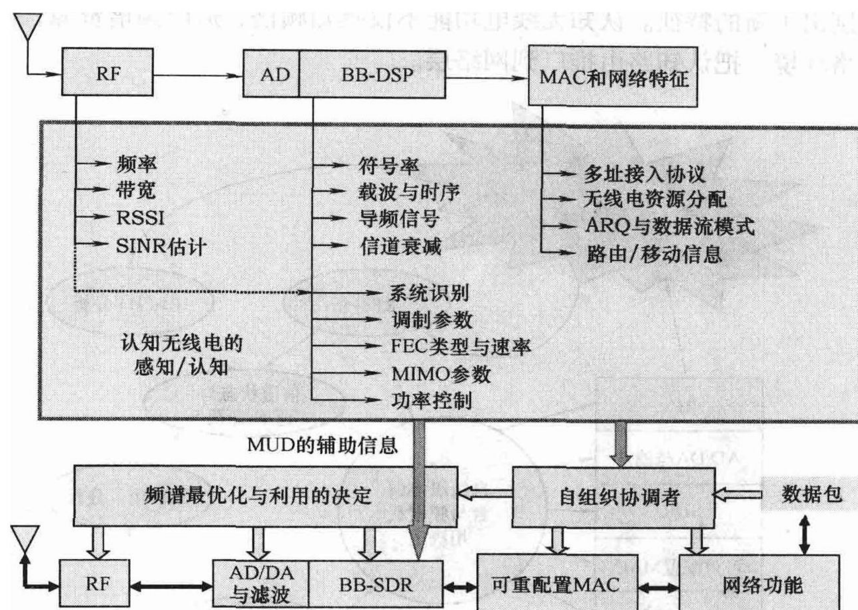


图 6-13　认知无线电网络终端的硬件结构

众所周知的是认知无线电是以频谱感知为核心的。可是，如图 6-13 所示，我们需要更多的信息来在实践中提高感知的性能，不仅限于频谱的感知，而且还需要对一些联网功能进行感知，成为广义的感知（或者认知）。我们把这样的频谱/网络感知的特点列在下面：

1）射频信号处理包括载波频率、信号带宽、信号强度（RSSI）、SINR 估计；

2）BB 预检测信号处理包括发射速率、载波和时序信息、导频信号、信道衰减；

3）BB 检测后处理（有的可以在检测前完成）包括系统/用户的审核、调制参数、纠错编码种类及速率、MIMO 参数、传输功率控制。

联网处理信息包括多址接入协议或者 MAC、无线电资源分配（例如时隙、子载波、码）、自动重复请求（ARQ）和流量模式（任意比特率（ABR）、恒定比特率（CBR）、可变比特率（VBR））、路由或者移动信息。以上列表的目的是实现频谱感知、检测共存的系统和网络，以及这些系统和网络的运行。研究文献提供了大量的实例来实现这个列表中的部分功能。认知无线电循环的工作流程如图 6-14 所示。

J. Mitola 在参考文献［1］和 Haykins 在参考文献［40］中发展了不同但是

类似的认知循环的概念。既然我们推广到认知无线网络，以及速率-距离概念，这就区别出了新的特征。认知无线电功能不仅感知频谱、适应频谱资源，同时也适应网络环境，把认知路由推广到网络层。

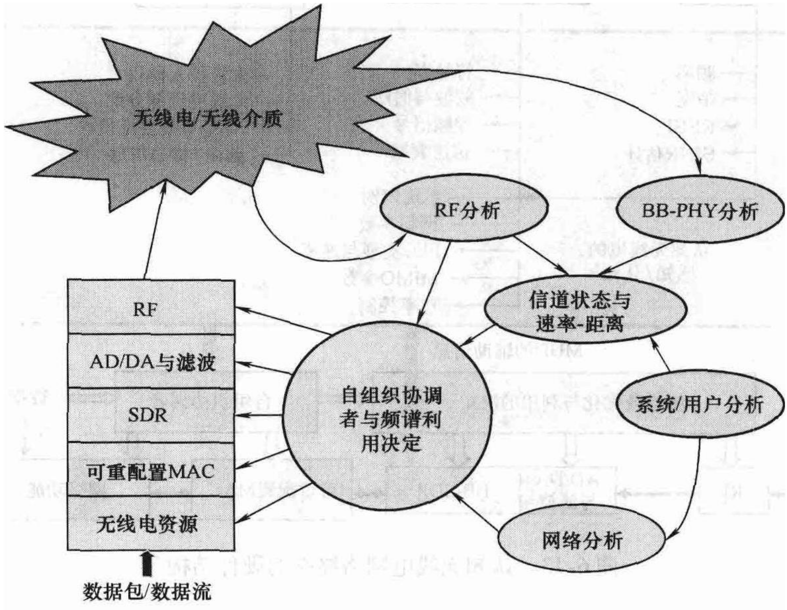


图 6-14 认知无线电网络终端的认知循环

6.3.2 可重配置 MAC

用于移动/泛在的计算的无线网络的媒体接入控制（MAC）是除了无线电收发机外的一个基本问题。Chen 在文献［4］中描述了一些衰落信道里的无线网络的基本挑战，并且建议了一些解决问题的原则。在一番努力后，一个统一的 MAC 算法在文献［5］中提出，以实现最知名的接入协议。它借助了 R. Gallager 的概念，即多用户接入实现载波监听（被推广为碰撞避免）或者实现碰撞检测，来形成 CATE 和 CRTE（碰撞避免/解决树结构来统一所有的协议）。以下算法和表 6-2 通过对相应的参数进行调整，以得到合适的多用户接入协议，更多的讨论参见第 8 章。

可重配置 MAC 算法

```
RP_1
if (access method = blocked) {
  allow new arrivals during previous cycle- > DN; }
if (memoryless_after_lost is set) {
```

```
  have all noted in DN call CATE (type_CATE;)}
else {
  unmarked nodes in DN call CATE (type_CATE);
  associate marked nodes in DN to group
    number # (original group number-g);}
  unmarked nodes in CN call CRTE (type_CRTE);
  associate marked nodes in CN to group
    number # (original group numbers-g);
  if (report grouping result is set) {
    all nodes report the grouping result back;}
set g=1; //start to process each group
RP_2
if (access method=free) {
  nodes with new arrival packets during the
  processing of group # (g-1) ->TX (g);}
nodes in group # g->TX (g);
process group # g with GP (gp_scheme);
if (there is no transmission) {
  g++;
if (G is set) {//G is the maximum TE size
      if (g>G) {goto RP_1;}
      else {goto RP_2;}}
    else {goto RP_2;}}
elseif (there is transmission) {
  if (access method=free) {
    nodes in group # (g+1)
      to group # (g+t) ->DN;}
      //t is the duration of the transmission
      If (the transmission is a success) {
        the successful node removes the
          transmitted packet from buffer;
      if (completeness is set) {
        g++;

      if (g>G) {goto RP_1;}
```

```
        else {goto RP_2;}}
      else {
        if memoryless_after_lost is set) {
        mark the loser in CN;}
      else {mark the loser in CN and DN;}
          current cycle ends and goto RP_1;}}
      else {
          collided nodes- >CN;
            if (completeness is set) {
              g++;
              if (g>G) {goto RP_1;}
            else {goto RP_2;}}
    else {
      if (memoryless_after_lost is set) {
       mark the loser in CN;}
      else {mark the loser in CN and DN;}
        current cycle ends and goto RP_1;}}}
```

6.3.3 无线电接入网络选择

遵循以上的体系结构，自组织协调者规划路由中正确的网络功能以控制QoS，并决定合适的MAC设置、软件无线电通信参数和射频参数。实现自组织无线通信的典型方法研究整个网络/系统的拓扑控制，并基于不同的准则进行优化。为了实际的实现，我们从另外一个角度考虑这个问题，即一个终端决定其路由，只取决于其拥有的信息，以选择正确的接入网络，包括认知无线电。无线接入网络可以是蜂窝网络的一部分，例如UTRAN，一个连接因特网的无线局域网的接入点，或者WiMAX系统的基站（或者是网状网络的一个用户站）。因此，我们假设有来自K个系统的用户，都在同一地理区域内工作，设备可以接入所有的工作频段。传统意义上，一个可以在某个系统内工作的移动设备无法在另外一个系统内工作，在这K个系统中的资源无法公平分配，有的系统可能拥挤，有的系统可能没有或者有很少的数据流。通过认知无线电，我们可以利用可能的协作，来提供个体和整体的性能。主要的挑战来自于在不同组合的系统中决定合适的协作以提供性能或者QoS。请注意在实际应用中，用户也许需要一个损耗函数作为系统性能指标。不失一般性，我们考虑一个有n_1个用户的电路交换（CS）网络（例如2G或者3G蜂窝网络），以及一个有n_2个用户的数据包交换（PS）的网络（例如WiFi）。对于这$N=n_1+n_2$个在两个系统间以认知模式工作

的用户，我们希望证明有效的路由能提高整体网络的性能（见图 6-15）。数据包的损失是因为冲突与重传，我们也假设网络中的用户数量相对比较稳定。

表 6-2 MAC 重配置的参数

	随机退后等待的 ALOHA	基本多进制 CRA	P-坚持 CSMA	CSMA/CA	GRAP
时隙	一次传输 + 一次反馈	一次传输 + 一次反馈	单传播延时	在规范中定义	单传播延时
接入方式	Free	Free	Free	Free	Blocked
完整性	No	No	No	No	Yes
丢失后内存是否减少	No	Yes	No	Yes	No
报告编组结果	No	No	No	No	Yes
批处理方法	双向握手	双向握手	双向握手	四向握手	阻止
CATE 类型	None	None	几何 CATE	BEB. CATE	均匀 CATE
CRTE 类型	几何 CRTE or BEB. CRTE	多进制 CRA. CRTE	几何 CRTE	BEB. CRTE	均匀 CRTE

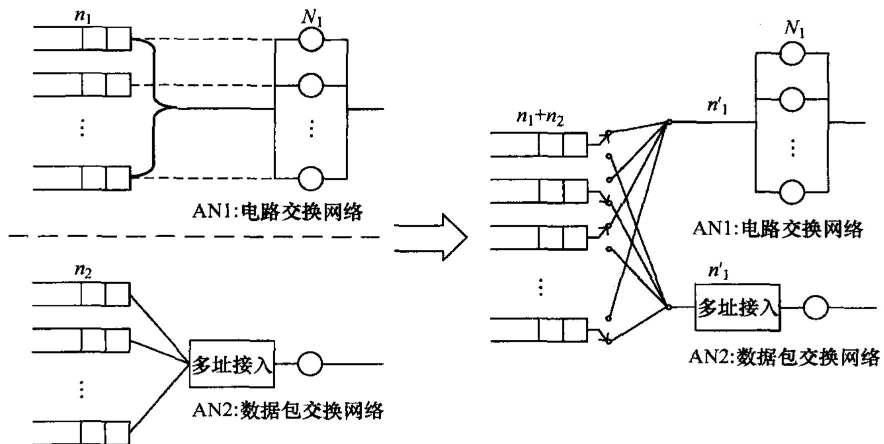


图 6-15 协作接入网络的队列模型

接入网络 1：电路交换网络被考虑成一个有 N_1 个服务器的多窗口队列模型。这些服务器的服务速度都是恒定值且等于 u_1。一个特定的用户只由一个服务器服务，不多于 N_1 个用户可以被接纳到网络中来。

接入网络 2：数据包交换网络被考虑为一个单服务器队列模型，在服务器前有一个多址接入设备，负责决定哪个用户获得接入的权利。这台服务器的服务速度是恒定的，等于 u_2，多址接入方式假设为时隙 ALOHA，重传概率为 q。

基于排队技术分析，终端可以做出对接入网络的正确选择，来为其数据流完成最优的路由。当然，这只是用来论证的最简单的情况。更一般的情况仍然是未解决的问题。

6.4 QoS预留分集无线电接入网络

为了扩展网络覆盖面、提供可靠性和增加数据传输率，不同的通信节点之间的协作/协同成为未来无线电网络（例如认知无线电网络）中新的设计范式。我们定义节点间的协作和协同如下：

多个节点间的协作意味着多个节点共同完成各自的目标，而多个节点间的协同意味着它们共同完成同一目标。

我们进一步将协作和协同划分为对称和不对称的情况。对称的协作/协同意味着节点互相帮助，而不对称协作/协同则意味着一个节点（网络）帮助其他节点（网络），但是其他节点（网络）并不帮助该节点（网络）。图6-16描述了分集网络的一般性网络拓扑。对该图我们解释如下：

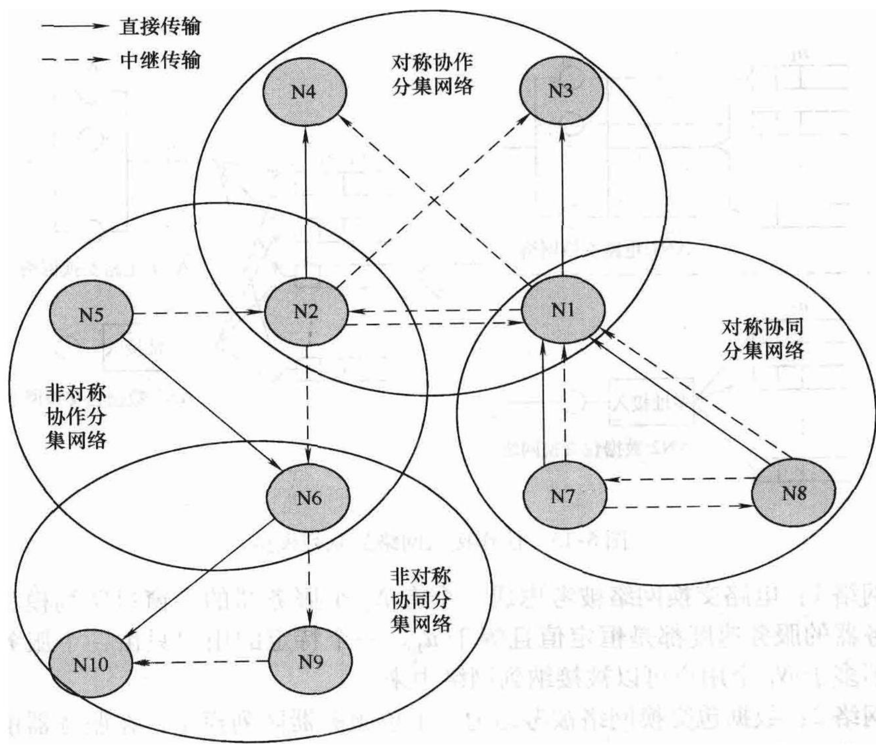


图6-16 广义分集网络拓扑

N1 和 N2 分别传输数据流至 N3 和 N4。N1 和 N2 互相帮助将数据流中继到终点。因此 N1、N2、N3 和 N4 形成对称协作分集网络。

N5 有数据流传至 N6，其数据流由 N2 完成中继。既然 N2 传输到 N4 的数据流并未经过 N5 的中继，所以 N2、N5、N6 就形成了非对称协作分集网络。

N7 和 N8 一起传输数据流至 N1，因此 N1、N7 和 N8 形成对称协同分集网络。

N6 将数据流传至 N10、N9 对其数据流进行中继，可是 N9 自身并无数据流可传。所以，N6、N9 和 N10 形成不对称协同分集网络。

对于这些定义，不同分集网络的情况已经被提出了，除了非对称协作分集网络的情况。

保证无线通信网络的 QoS 是一个永恒的挑战[34-37]。主要的原因是无线信道容量是时变和突发式的，这使得数据率对信道变化敏感。一旦信源速率超过系统提供的数据率，数据包的到达时间就可能超过最大容忍值，这个包在延时敏感的数据流里就会失效。在用户协作的原始概念里，Sendornaris、Erkip 和 Aazhang 指出协作/协同中继会减轻数据率对信道变化的敏感度，这就是对 QoS 的最有力保证，甚至当协作/协同无法带来其他好处（例如数据率的增加）时。这是因为一些实时的应用（例如音频或者视频）有最小数据率的要求，协作/协同则可以降低中断的概率，从而改善 QoS。可是，通过协作/协同中继来提供 QoS 的方法很少有人研究。本节我们将回顾一些分集网络上的 QoS 的现存工作。基于这些工作，我们将指出有关提供 QoS 的一些尚未解决的问题。

6.4.1　协作/协同分集和高效协议

一个基本的协作/协同网络拓扑包括一个源节点、一个中继节点和一个目的节点，如图 6-17 所示。以下是中继节点的工作方式：

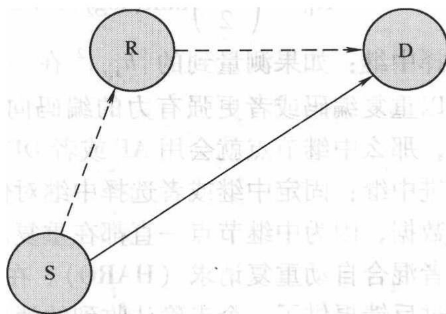


图 6-17　基本协作/协同网络拓扑

提供方便：中继节点不直接主动帮助源节点，只是通过尽量少产生干扰来为源节点提供方便。

协作/协同：1）中继节点完全译码并重传（即译码转发，DF）或者 2）中继节点只放大原信息并重传（即放大转发，AF）。

观测：中继节点将接收到的量化后的信号进行编码并重传（即压缩转发，CF）。

既然全工方式的接收机的设计是复杂的，那么消除节点间的干扰可以极大地简化接收机的设计。为了保证半工式工作，信道可以被划分成正交的子信道，如

图6-18所示。为了提供方便，时分信道分配可以采用图6-18a、b的方式；否则，时分信道分配如图6-18c所示。

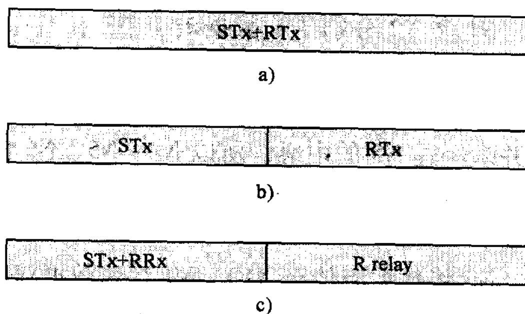


图6-18 a）源节点和中继节点同时传输
b）源节点先传输，然后中继节点再传输
c）源节点将数据包传给中继节点，然后中继节点将数据包传给目的节点

如第4章所描述的，现存的中继协议可以划分为以下三种类型：

固定中继：中继节点要么根据其发射功率的限制来放大接收到的信号，要么解码、重编码并重传接收到的信息。

放大转发：假设加性噪声被建模为独立零均值循环对称复白高斯噪声，并且在每个接收机处的方差为1，最大互信息，也就是可以达到的最大传输速率为

$$R_{\mathrm{AF}}=\left(\frac{T_fB}{2}\right)\log_2\left(1+2\gamma_1P_s+\frac{4\gamma_2P_s\gamma_3P_c}{1+2\gamma_2P_s+2\gamma_3P_c}\right)\tag{6-31}$$

式中，B为系统带宽，P_s和P_c分别为源节点和中继节点的平均发射功率，T_f为帧长，$\gamma_1=|h_{s,r}|^2$，$\gamma_2=|h_{s,d}|^2$和$\gamma_3=|h_{r,d}|^2$为源节点-中继节点，源节点-目的节点和中继节点-目的节点的信道增益。

译码转发：在译码转发模式中，中继节点如果能够译码成功，就将信息转发给目的节点；其最大互信息和最大传输速率为

$$R_{\mathrm{DF}}=\left(\frac{T_fB}{2}\right)\min(\log_2(1+2\gamma_2P_s),\ \log_2(1+2\gamma_1P_s+2\gamma_3P_r))\tag{6-32}$$

选择中继：如果测量到的$|h_{s,r}|^2$在一定阈值以下，那么源节点就可以简单地继续以重复编码或者更强有力的编码向目标发射。如果测量到的$|h_{s,r}|^2$在阈值以上，那么中继节点就会用AF或者DF进行转发，以期获得分集增益。

渐进中继：固定中继或者选择中继对信道自由度的使用是低效的，尤其对于高速率数据，因为中继节点一直都在重复。渐进中继协议可以看成渐进冗余的推广，或者混合自动重复请求（HARQ）。在ARQ里，源节点重新传输，如果目的节点通过反馈提供了一个未确认收到的消息。在渐进中继里，中继节点重传以求利用空间分集。

6.4.2 无线非对称协同中继网络的统计QoS保证

实时多媒体服务，例如视频和音频，需要有限延时，或者说是保证的带宽。一旦接收到的实时数据报违反了延时的要求，它就会被认为是无效的而被抛弃。

可是，在无线移动网络里，一个硬性的延时要求保证实际上是无法实现的，因为时变衰落信道的缘故。例如在瑞利衰落信道上，唯一的系统带宽的确定性下界就是0。因此，我们应该考虑一个别的方法来提供统计 QoS 保证，即以较小的违反概率来保证延时的限制。为此，Wu 和 Negi 在参考文献［33］中发展了有效容量的概念，这个概念来自于统计 QoS 保证的有效带宽定理。除了在一般无线通信里提供统计 QoS 保证外，有效容量也为中继网络提供了 QoS 的新的设计方法。在本节里我们回顾有效带宽和有效容量的理论，并将其应用于非对称协同中继网络，以期实现延时 QoS 的要求。

6.4.2.1　有效带宽和有效容量

在20世纪90年代，有效带宽理论的领域发展出了统计 QoS 保证的框架。在相关文献里，研究者们使用参数 θ 来表示 QoS 保证的程度。特别地，C.-S. Chang 在参考文献［36］中证明了对于一个有稳态遍历到达和服务过程的排队系统，其队列长度过程 $Q(t)$ 收敛到一个随机变量 $Q(\infty)$，使得

$$-\lim_{x\to\infty}\frac{\log\ (Pr\{Q(\infty)>x\})}{x}=\theta \tag{6-33}$$

以上定理说明了队列长度超过某一个数值 x 的概率随 x 的增长而呈指数下降。参考文献［1］中的参数 $\theta(\theta>0)$ 代表了违反 QoS 的概率呈指数下降的速度。一个小的 θ 意味着缓慢的下降速度，也就是说系统提供了较为松散的 QoS 保证。另一方面，大的 θ 对应于较快的下降速度，因此系统可以提供较为严格的 QoS 保证。

从有效带宽理论出发，Wu 和 Negi 在参考文献［33］中提出了有效容量的概念，其定义如下：

一个为保证用 θ 给定的 QoS 要求，给定的服务速率所能支持的最大恒定到达速率

其表达式为

$$E_c(\theta)=-\frac{1}{\theta}\log(E[\mathrm{e}^{-\theta R}]) \tag{6-34}$$

式中，R 是独立同分布的服务过程，$\mathrm{E}[Y]$ 是 Y 的期望。特别的，如果 $\theta\geqslant -\log\varepsilon/D_{\max}$，就有

$$\sup_t Pr\{D(t)\geqslant D_{\max}\}\leqslant\varepsilon \tag{6-35}$$

式中，$D_{\max}$是数据流的最大容忍延时，$D\ (t)$ 是时间 t 的延时。参考文献［3］说明了数据流超过最大可容忍延时的概率低于 ε。

6.4.2.2　给定 QoS 约束时的 AF 和 DF 中继的动态资源分配

跨层无线电资源分配的方案可以基于有效容量，用于非对称协同中继网络，以满足无线多媒体通信延时 QoS 的要求。为了在给定 QoS 约束下用 θ 表示最大

化有效容量，问题被建模成一个源功率和中继功率分配的问题。因此，AF 中继的最大化问题可以表达为

$$\arg\max_{P(\gamma)}\left\{-\frac{1}{\theta}\log(E_\gamma[\exp(-\theta R_{\mathrm{AF}}(\gamma))])\right\} \tag{6-36}$$

约束于以下功率限制

$$\begin{cases}\mathrm{E}_\gamma[P_s(\gamma)+P_r(\gamma)]\leqslant\overline{P}\\ P_s(\gamma)\geqslant 0 \text{ 且 } P_r(\gamma)\geqslant 0\end{cases} \tag{6-37}$$

式中，$P(\gamma)\triangleq(P_s(\gamma),\ P_r(\gamma))$，$\overline{P}$ 是平均功率限制。

最优的功率分配策略为

$$\begin{cases}P_s(\gamma)=uP_r(\gamma)\\ P_r(\gamma)=\frac{1}{v}\left(\left[\left(\frac{\gamma_0}{\gamma_3}\right)\left(\frac{\gamma_3+c}{\gamma_1+c}\right)^2\right]^{-\frac{2}{\beta+2}}-1\right)\end{cases} \tag{6-38}$$

式中，$c=\sqrt{\gamma_1\gamma_3+\gamma_2\gamma_3-\gamma_1\gamma_2}$，$\gamma_0$ 为切断阈值，由平均网络功率限制所决定，$\beta\triangleq\theta T_f B/\log2$，如果 $P_s(\gamma)>0$，$P_r(\gamma)>0$，就

$$\begin{cases}u=\frac{\gamma_3(\gamma_1+c)}{(\gamma_3-\gamma_1)\gamma_2}\\ v=\frac{2c\gamma_3(\gamma_1+c)^2}{(\gamma_3-\gamma_1)\gamma_2(\gamma_3+c)}\end{cases} \tag{6-39}$$

否则，该策略退化为直接传输，P（γ）为

$$\begin{cases}P_s(\gamma)=\frac{1}{2}\left[\left(\gamma_0^{\frac{2}{\beta+2}}\gamma_1^{\frac{\beta}{\beta+2}}\right)^{-1}-\gamma_1^{-1}\right]^+\\ P_c(\gamma)=0\end{cases} \tag{6-40}$$

AF 中继的最大化问题可以表达为

$$\arg\max_{P(\gamma)}\left\{-\frac{1}{\theta}\log(E_\gamma[\exp(-\theta R_{\mathrm{DF}}(\gamma))])\right\} \tag{6-41}$$

最优的功率分配为

$$\begin{cases}P_s(\gamma)=\frac{1}{2}\left[\left(\gamma_0^{\frac{2}{\beta+2}}\gamma_1^{\frac{\beta}{\beta+2}}\right)^{-1}-\gamma_2^{-1}\right]^+\\ P_c(\gamma)=0\end{cases} \tag{6-42}$$

如上所示，Tang 和 Zhang 在参考文献［38］中通过将这个最优化问题建模为功率分配问题，最大化了系统可以支持的到达速率，并统计地满足延时的要求，这个解决方法主要是用于非对称的协同式中继，其中中继节点并无数据流可传。尽管有效容量可以在不用非对称中继进行直接传输时保证统计 QoS 要求，它在应用于对称协同中继（其中的中继节点和源节点都有数据包可传）时过于复杂。

6.4.2.3 OFDMA 非对称协同中继

除普通的物理传输外，人们开始对非对称的正交频分多址（OFMDA）上的协同式中继产生兴趣，因为用于多跳中继网络的 IEEE 802.16j 的被采用。可是，因为载波间干扰消除和 OFDMA 的信道估计仍然有困难，特别是对于上行链路，协同中继现在主要是应用在下行链路上。Pischella 和 Belfiore 在参考文献［26］中提出了一种 OFDMA 下行协作情况，其中两个基站（BS）用非正交 AF 协作协议来提供并维护移动站（MS）的 QoS。在这个下行协同框架里，两个基站使用同样的 FFT 规模 N_{FFT}，总带宽为 B。每一个移动终端 k，由其原始的基站（$BS_{s,k}$）服务，并可能被其最近的基站（$BS_{r,k}$）中继。QoS 驱动的资源分配问题是指每一个基站致力于为有性能保障（GP）的用户提供目标数据率，并最大化最大努力（BE）的用户的速率和。因此，无线电资源分配问题可以建模为以下最优化问题：

$$\min \sum_{k=1}^{K_1} \sum_{m=1}^{n_{SC}} c_{k,m} P_{BS,k,m}$$

所以

$$D_k = D_{target,k}, \quad k \in [1, K_1] \tag{6-43}$$

$$\max \sum_{k=K_1+1}^{K_1+K_2} D_k$$

所以

$$\sum_{k=1}^{K_1+K_2} \sum_{m=1}^{n_{SC}} c_{k,m} P_{BS,k,m} \leqslant P_{max}$$

式中，m 为子载波的序号，P_{max}为最大功率限制，$P_{s,k,m}(P_{r,k,m})$ 为从源节点（中继节点）到用户 k 在用户 m 处的发射功率。D_k 为用户 k 的数据速率，$c_{k,m}$是用户 k 在第 m 个子载波上的功率分配的指标，且 $c_{k,m} \in [0, 1]$。前 K_1 个用户为 GP 用户，接下来的 K_2 个用户为 BE 用户。GP 用户可以是直接或者中继用户。

以上提及的分集网络主要是非对称协同的场景。如何将协作中继应用到认知无线电的 QoS 保证，还是一个公认未解决的问题。

6.5 Ad hoc 和认知无线电网络的标度律

既然认知无线电网络允许多个传输，认知无线电网络的总速率也许可以更好地代表其性能，如 M. Vu、N. Devroye、V. Tarokh 在参考文献［22］中所建议。（无线）网络的标度律一般用来建立总速率的上界或者下界。下界可以通过一定的传输策略获得，因为这个传输速率可以提供可行的速率。上界可以由理论获得。如果下界和上界相等，我们就获得了准确的总速率的标度律。

6.5.1 网络和信道模型

假定一个平面上有 n 对无线设备（或者用户、或者节点）相互通信。每一

对节点都包括一个发射机和一个接收机。我们可以有两种网络模型：密集网络和广域网络。在密集网络中，网络的区域是有限的，而节点数量在增加。在广域网络中，节点的密度是保持不变的，这意味着网络面积和节点数呈线性增长。

网络性能的指标为总的网络容量，即总速率或吞吐量，其定义为

$$C(n)=\sum_{i=1}^{n}R_i$$

式中，R_i 为通信对 i 的信息速率。每一个通信对的容量定义为

$$R(n)=\frac{C(n)}{n}$$

我们要知道当 n 趋向无穷时，容量是如何增长的。我们也将传播距离、阴影效应和衰落考虑进信道模型。给定发射机和接收机间的距离，信道增益为

$$h(d)=\frac{e^{-\gamma d}}{d^{\alpha/2}}h_s$$

式中，d 为发射机和接收机之间的距离，α 为路径功率损失的指数，γ 为吸收/延时的常数，h_s 为阴影效应衰减因子。为了简化分析，我们令 $h_s=1$ 和 $\gamma=0$，只考虑路径损失，则当 $\alpha\geqslant2$ 时是合理的[22]。

6.5.2 Ad hoc 网络

我们考虑节点随机均匀分布的 Ad hoc 网络，发射机-接收机对也是随机的。在 $\alpha\geqslant2$ 时，每一个传输对需要最小的接收 SINR 为

$$\frac{P/d_i^{\alpha}}{\sigma_n^2+\sum_{i\neq k}P/d_{ik}^{\alpha}}\geqslant\beta$$

式中，α_n^2 为噪声功率，d_k 为第 k 对发射机-接收机的距离，d_{ik}为接收节点 k 和干扰节点 i 的距离。每个节点的传输速率的下界为

$$R(n)\geqslant\frac{c(\beta)}{\sqrt{n\log n}}$$

式中，$c(\beta)\propto\beta^{-1/2\alpha}$为常数，在多跳路由里可以用最近邻居转发的方法实现。其上界可以在没有干扰限制时获得

$$R(n)<\frac{c_1}{\sqrt{n}}$$

式中，c_1 是一个常数。我们可以证明在广域网络里，每个节点的流量为 $1/\sqrt{n}$[24]。

当节点呈泊松点过程分布时，每个节点吞吐量的上限为 $1/\sqrt{n}$，这可以由渗流理论获得。

如果我们通过物理层处理进一步允许节点间协作的话，每个节点的吞吐量可

以渐近为常数，其方法为分层聚簇，在簇内的通信为 Ad hoc 的，而在簇间的通信则采用 MIMO 通信。对于广域网络，当 $2<\alpha\leqslant 3$ 时，

$$C(n)\sim n^{2-\alpha/2}$$

当 $\alpha>3$ 时，最近邻居多跳方法为最优的，而且有

$$C(n)\sim\sqrt{n}$$

6.5.3 认知无线电网络

我们现在研究认知无线电网络，其中有两种节点，主系统和认知无线电（例如频谱次要用户），如图 6-19 所示。我们再次考虑广域网络，主系统的发射机（PS-Tx）在中心，其频谱占有半径为 R_0，主系统接收机（PS-Rx）在此半径内。认知无线随机均匀分布在主系统频谱占有半径外，其密度为 λ。认知通信以单跳形式存在，最大传输距离为 $D_{\max}$。在这个方案中，任意的干扰发射机都有非零距离。

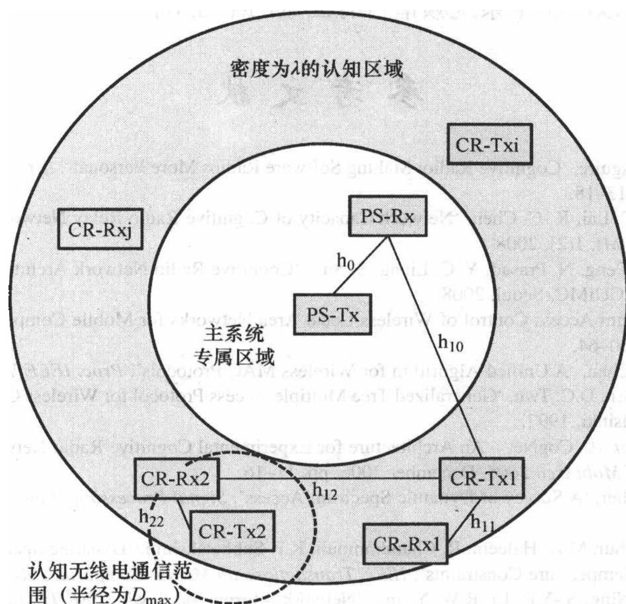


图 6-19 认知无线电网络设置

我们引入失效概率这一概念，即主系统接收到的信号低于某一指标。假定主系统发射机以功率 P_0 发射，每个认知无线电发射机以功率 P_c 发射。没有多用户检测时，每一个接收机将其他传输（其实就是干扰）看成噪声。当 $\alpha>2$ 时，认知无线电的平均速率和和节点数 n 呈线性增长。每个节点的速率以很高的概率保持常数，以保证

$$R(n) \geq \log\left(1 + \frac{P_{\min}}{\sigma_{\max}^2 + I_{\text{worst}}}\right)$$

式中，$P_{\min} = \frac{P_c}{D_{\max}^{\alpha}}$，$\sigma_{\max}^2 = \sigma_n^2 + P_0/R_0^{\alpha}$，$\sigma_n^2$ 是噪声功率。最糟糕的干扰就是 I_{worst}，

$$I_{\text{worst}} = \frac{2\pi\lambda P_c}{(\alpha - 2)\varepsilon^{\alpha - 2}}$$

$\varepsilon > 0$ 在每个节点获得常数吞吐量里起了关键作用，这等价于每个认知无线节点周围有一个频谱独占区域，其他认知无线节点无法在其中运行。最糟糕的情况里，发射机和接收机的距离可以$\sqrt{n}$增长，致使每个节点的吞吐量下降到 $1/\sqrt{n}$。

以 C_0 为主系统发射机的传输速率，η 为给定的阈值，其最坏情况下中断的约束为

$$P[C_0 \leq \eta] \leq b$$

中断来自于认知无线电节点的随机部署。通过给主系统接收机的平均干扰设置上限，我们可以限制主系统频谱占有区域的半径 R_0。

参考文献

[1] J. Mitola, G.Q. Maguire, 'Cognitive Radio: Making Software Radios More Personal', *IEEE Personal Communications*, **6**(4), 1999, 13–18.

[2] C.-H. Huang, Y.-C. Lai, K.-C. Chen, 'Network Capacity of Cognitive Radio Relay Networks', *Physical Communications (PHYCOM)*, **1**(2), 2008.

[3] K.-C. Chen, Y.-J. Peng, N. Prasad, Y.-C. Liang, S. Sun, 'Cognitive Radio Network Architecture: Part I – General Structure', ACM ICUIMC, Seoul, 2008.

[4] K.C. Chen, 'Medium Access Control of Wireless Local Area Networks for Mobile Computing', *IEEE Networks*, September 1994, 50–64.

[5] C.M. Teng, K.C. Chen, 'A Unified Algorithm for Wireless MAC Protocols', *Proc. IEEE VTC*, 2002.

[6] Y.K. Sun, K.C. Chen, D.C. Twu, 'Generalized Tree Multiple Access Protocol for Wireless Communications', *Proc. IEEE PIMRC*, Helsinki, 1997.

[7] D. Raychaudhuri *et al.* 'CogNet – An Architecture for Experimental Cognitive Radio Networks within the Future Internet', *Proc. of MobiArch 2006*, December 2006, pp. 11–16.

[8] Z. Qing, B.M. Sadler, 'A Survey of Dynamic Spectrum Access', *Signal Processing Magazine, IEEE*, **24**(3), 2007, 79–89.

[9] Y. Xing, C.N. Mathur, M.A. Haleem, R. Chandramouli, K.P. Subbalakshmi, 'Dynamic Spectrum Access with QoS and Interference Temperature Constraints', *IEEE Transactions on Mobile Computing*, **6**(4), 2007, 423–433.

[10] R. Ahlswede, C. Ning, S.-Y.R. Li, R.W. Yeung, 'Network information flow', *IEEE Transactions on Information Theory*, **46**(4), 2000, 1204–1216.

[11] L. Song; R.W. Yeung, N. Cai, 'Zero-error Network Coding for Acyclic Networks', *IEEE Transactions on Information Theory*, **49**(12), 2003, 3129–3139.

[12] L. Geng, Y.-C. Liang, F. Chin, 'Network Coding for Wireless Ad Hoc Cognitive Radio Networks', *Personal, Indoor and Mobile Radio Communications, 2007 (PIMRC 2007), IEEE 18th International Symposium*, September 2007, pp. 1–7.

[13] L. Xiao; T. Fuja, J. Kliewer, D. Costello, 'A Network Coding Approach to Cooperative Diversity', *IEEE Transactions on Information Theory*, **53**(10), 2007, 3714–3722.

[14] S. Fu, K. Lu, Y. Qian, M. Varanasi, 'Cooperative Network Coding for Wireless Ad-Hoc Networks', *Global Telecommunications Conference, 2007 (GLOBECOM'07), IEEE*, 26–30 November 2007, pp. 812–816.

[15] J. Mitola, 'Cognitive Radio for Flexible Mobile Multimedia Communications', *1999 IEEE International Workshop on Mobile Multimedia Communications (MoMuC '99)*, 1999, pp. 3–10.

[16] L.R. Ford, Jr., D.R. Fulkerson, *Flows in Networks*, Princeton University Press, New Jersey, 1962.
[17] T.C. Hu, 'Multicommodity Network Flows', *Operations Research*, **11**,1963, 344–360.
[18] R.E. Gomory, T.C. Hu, 'Synthesis of a Communication Network', *J. SIAM*, **12**,1964, 348–369.
[19] R. Hassin, 'On Multicommodity Flows in Planar Graphs', *Networks*, **14**(2), 1984, 225–235.
[20] T.H. Cormen, C.E. Leiserson, R.L. Rivest, C. Stein, *Introduction to Algorithms*, The MIT Press, Cambridge, MA, 2001, 788–789.
[21] K.C. Chen, *et al.* 'Terminal Architecture for Cognitive Radio Networking Devices', *Proceeding Wireless Personal Multimedia Communication Conference*, 2007.
[22] M. Vu, N. Devroye, V. Tarokh, 'An Overview of Scaling Laws in Ad-hoc and Cognitive Radio Networks', *Wireless Personal Communications*, **45**,2008, 343–354.
[23] P. Gupta, P.R. Kumar, 'The Capacity of Wireless Networks', *IEEE Transactions on Information Theory*, **2**(46), 2000, 388–404.
[24] P. Tarasak, Y.H. Lee, 'Joint Cooperative Diversity and Scheduling in OFDMA Relay Systems', *IEEE Wireless Communications and Networking Conference (WCNC)*, 2008.
[25] M. Pischella, J.C. Belfiore, 'QoS-Based Resource Allocation with Cooperative Diversity in OFDMA', *IEEE Vehicular Technology Conference (VTC)*, 2008.
[26] M. Pischella, J.C. Belfiore, 'Power Control in Distributed Cooperative OFDMA Cellular Networks', *IEEE Trans. Wireless Commun.*, **7**(5), 2008, 1900–1906.
[27] T.Q.S. Quek, H. Shin, M.Z. Win, 'Robust Wireless Relay Networks: Slow Power Allocation with Guarantee QoS', *IEEE Journal of Selected Topics in Signal Processing*, **1**(4), 2007, 700–713.
[28] S.-J. Kim, X. Wang, M. Madihian, 'Optimal Resource Allocation in Multi-hop OFDMA Wireless Networks with Cooperative Relay', *IEEE Trans. Wireless Commun.*, **7**(5), 2008, 1833–1838.
[29] A. Fallahi, E. Hossain, A.S. Alfa, 'QoS and Energy Trade Off in Distributed Energy-limited Mash/Relay Networks: A Queuing Analysis', *IEEE Trans. Parallel and Distributed Systems*, **17**(6), 2006, 576–592.
[30] S. Ren, K.B. Letaief, 'Optimal Effective Capacity for Cooperative Relay Networks with QoS Guarantees,' *IEEE International Conference on Communications*, 2008.
[31] C. Sun, K.B. Letaief, 'User Cooperative in Heterogeneous Cognitive Radio Networks with Interference Reduction', *IEEE International Conference on Communications*, 2008.
[32] J.N. Laneman, D.N.C. Tse, G.W. Wornell, 'Cooperative Diversity in Wireless Networks: Efficient Protocols and Outage Behavior', *IEEE Trans. Information Theory*, **50**(12), 2004, 3062–3079.
[33] D. Wu, R. Negi, 'Effective Capacity: A Wireless Link Model for Support of Quality of Service', *IEEE Trans. Wireless Commun.*, **2**(4), 2003, 630–643.
[34] D. Wu, 'Providing Quality of Service Guarantees in Wireless Networks,' Ph.D Dissertation, Carnegie Mellon University, PA, 2003.
[35] D. Wu, R. Negi, 'Downlink Scheduling in a Cellular Network for Quality of Service Assurance', *IEEE Trans. Veh. Technol.*, **53**(5), 2004, 1547–1557.
[36] C.-S. Chang, 'Stability, Queue Length, and Delay of Deterministic and Stochastic Queuing Networks', *IEEE Trans. Automat. Contr.*, **39**(5), 1994, 913–931.
[37] A.I. Elwald, D. Mitra, 'Effective Bandwidth of General Markovian Traffic Sources and Admission Control of High Speed Networks', *IEEE/ACM Trans. Networking*, **1**(3), 1993, 329–341.
[38] J. Tang, X. Zhang, 'Cross-layer Resource Allocation Over Wireless Relay Networks for Quality of Service Provisioning', *IEEE Journal of Selected Areas in Communications*, **25**(4), 2007, 645–656.
[39] R.U. Nabar, H. Bölcskei, F.W. Kneubühler, 'Fading Relay Channel: Performance Limits and Space-time Signal Design', *IEEE Journal of Selected Areas in Communications*, **22**,2004, 1099–1109.
[40] S. Haykin, 'Cognitive Radio: Brain-empowered Wireless Communications', *IEEE Journal on Selected Areas in Communications*, **23**(2), 2005, 201–220.

第7章　频谱感知

为了认知无线电的可靠运行，我们必须在链路层精确检测频谱空洞（也就是，在某一时刻没有被使用的频段），这就给予了频谱感知一个重要的角色，产生了大量的研究。对于叠加或者共存式的多无线电系统，我们也许想要感知更多的频谱信息，来控制干扰的大小。当利用认知无线电进行联网时，我们需要链路建立以外的更多信息，以在网络层获得更好的频谱利用。

7.1　频谱感知以检测特定的主系统

我们可以从最简单的情况开始频谱感知，即一个主系统（PS）的发射机-接收机对和一个次要认知无线电发射机-接收机对。除了有效的频谱感知外，我们还有更多的挑战。次要认知无线电节点，甚至是在系统的边缘或者警戒频段，应该可以检测到主信号，尽管有可能无法译码。进一步地，次要认知无线电节点一般来说不知道主系统节点的发射策略，而且可能无法知道主系统节点的训练和同步信号。这意味着次要用户被迫进行非相干能量检测，在低信噪比（SNR）时其性能比相干检测差得多。

7.1.1　传统的频谱感知

既然认知无线电经常被考虑成智能型无线通信设备，可以感知环境，通过调整自己来对传输和频谱利用进行最优化。频谱感知为认知无线电用户提供了重要功能，使之可以检测主系统未使用的频谱，提高整体频谱效率。在链路层针对单一主用户系统的频谱感知被用于在两个假设间做一个判决，即

$$y[n]=\begin{cases}w(n) & \\ hs[n] & +w[n]\end{cases} \quad n=1,\ \cdots,\ N \quad \begin{matrix}H_0\\ H_1\end{matrix} \tag{7-1}$$

式中，$y[n]$ 为认知无线电接收到的复信号，$s[n]$ 为主用户发射的信号，$w[n]$ 为加性高斯白噪声（AWGN），h 为理想信道的复增益，N 为观察间隔。如果信道不是理想的，h 和 $s[n]$ 就是卷积的而不是相乘的。H_0 代表零假设，即没有主用户，H_1 代表另外一个假设，即主用户信号存在。Wang 等将频谱感知分为基于能量的和基于特征的[4]。以下，我们概述一下一些著名的频谱感知的方法，即能量检测、匹配滤波器、周期稳态检测和小波检测（基于能量的）。

7.1.1.1　能量检测

当主用户的信号不被获知时，能量检测的方法是在所有检测零均值调制信号的方法里最优的，并且可以应用到认知无线电里。在能量检测方法中，无线电频率能量或者接收信号强度指示（RSSI）在一个观测时间内被测量，然后决定频谱是否被占用。能量检测经常应用在时域上，有时也可以在频域上实现。我们关注在时域实现上。接收到的信号被二次方后，在整个观测期间被积分。然后积分器的输出与一个门限值比较，用来决定主用户是否存在，即以下二元决策：

$$\begin{cases} H_0, & \sum_{n=1}^{N} |y[n]|^2 \leqslant \lambda \\ H_1, & \text{其他} \end{cases} \tag{7-2}$$

式中，λ 为门限值，取决于接收机噪声。

考虑到实现，有多种基于能量检测的传感器。模拟实现需要一个模拟的固定带宽滤波器，但是如果要同时感知窄带和宽带的信号，这个滤波器就缺乏灵活性。数字实现可以通过基于 FFT 的谱估计，更加灵活，这样的结构固有地支持各种带宽类型，并允许同时检测多个信号。图 7-1 所示为能量检测器数字实现的结构。

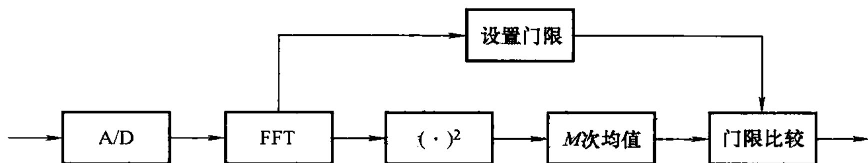


图 7-1　能量检测器的数字实现

尽管能量检测方法不需要主用户信号的先验知识，但是它仍然有各种困难。首先，它只能检测能量超过阈值的主用户。能量检测的阈值也可能有问题，因为它容易受变化的背景噪声，尤其是干扰的影响。另外一个问题是能量检测无法区分同一信道内的主用户信号和其他次要用户信号。当多个系统的主用户在认知无线电里共存时，这是一个严重的挑战，我们随后在本章里进行讨论。

7.1.1.2　匹配滤波器

匹配滤波器是最优的检测方法，因为它可以在加性高斯白噪声中最大化地接收到信号的信噪比。可是，匹配滤波器需要主用户信号在物理层和 MAC 层的先验信息，例如脉冲成形、调制方式、数据包格式。所以，匹配滤波器是一种基于特征的频谱感知。匹配滤波器是通过对一个已知信号和一个未知信号做相关来检测其存在。这等价于用时间反转的假设信号和未知信号进行卷积。然后匹配滤波器的输出和一个门限做比较，来判断主用户信号是否存在。因此，我们做以下的二元判决：

$$\begin{cases} H_0, & \sum_{n=1}^{N} y[n]x[n]^* \leqslant \lambda \\ H_1, & \text{其他} \end{cases} \tag{7-3}$$

式中，λ 为门限。

匹配滤波器一般应用在无线电通信和雷达传输上。可是，在认知无线电的范畴内，匹配滤波器的应用严重受到限制，如果主用户信号的信息很难得到的话。幸运的是，绝大多数有执照的系统有导频信号、前导、同步信号或者扩频码，以获得相关检测。如果主用户信号的部分信息已知，例如导频信号或者前导，匹配滤波器仍然可以使用。现以相关导频信号作为例子。这里，式（7-1）和式（7-3）中的 $x[n]$ 被 $x_p[n]$ 所代替，即主用户的导频信号。图 7-2 所示为相关导频检测器数字实现的结构。其主要好处是匹配滤波器需要更少的时间来得到高的处理增益，因为采用了相关检测。可是，对于解调，认知无线电不得不实现时间和载波的同步，甚至信道均衡。因此，认知无线电对每一种主系统都需要一个专门的接收机，这就增加了复杂度，当目标系统可能是多种系统之一时，这就对匹配滤波器甚至对于可编程的实现造成了严重的挑战。

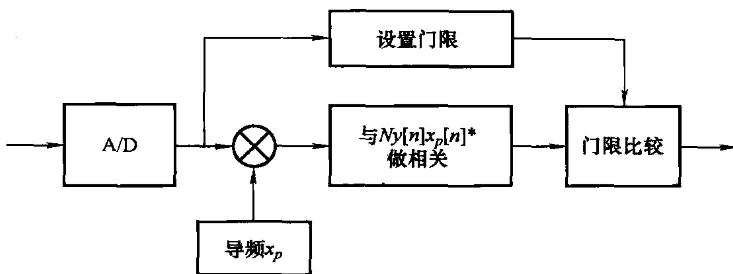


图 7-2 相关导频检测器的数字实现

7.1.1.3 周期稳态检测

周期稳态检测的想法是利用调制信号内在的周期性，例如正弦载波、脉冲序列、重复扩展、跳频序列或者循环前缀。一个信号被称为（宽）周期稳态，当其自相关函数是时间 t 的周期函数。周期稳态检测可以如下进行。首先，我们可以计算观测到的信号 $x(t)$ 的周期自相关函数（CAF）$R_x^\alpha(\tau)$：

$$R_x^\alpha(\tau) = \lim_{T\to\infty} \frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} x\left(t+\frac{\tau}{2}\right)x\left(t-\frac{\tau}{2}\right)\mathrm{e}^{-\mathrm{j}2\pi\alpha t}\mathrm{d}t \tag{7-4}$$

式中，α 称为周期频率。然后，我们可以计算 CAF 的离散傅里叶变换，来得到谱相关函数：

$$S_x^\alpha(f) = \int_{-\infty}^{\infty} R_x^\alpha(\tau)\mathrm{e}^{-\mathrm{j}2\pi f\tau}\mathrm{d}\tau \tag{7-5}$$

特别的，以下式（7-6）已经得到证明，

$$S_x^\alpha(f) = \lim_{T\to\infty}\lim_{Z\to\infty}\frac{1}{TZ}\int_{-\frac{Z}{2}}^{\frac{Z}{2}} X_T\left(t,f+\frac{\alpha}{2}\tau\right)X_T^*\left(t,f-\frac{\alpha}{2}\tau\right)\mathrm{d}t \tag{7-6}$$

其中，

$$X_T(t,f) = \int_{t-T/2}^{t+T/2} x(u)\,\mathrm{e}^{-\mathrm{j}2\pi fu}\mathrm{d}u \tag{7-7}$$

谱相关函数 $S_x^\alpha(f)$ 也称为周期频谱，它是频率和周期频率的二元函数。我们也要注意功率谱只是谱相关函数当 $\alpha=0$ 时的特例。最后，搜索对应于 SCF 平面上的尖峰的唯一周期频率，以完成检测任务。

循环稳态检测器一般是在数字域内实现的。直接的算法先通过 FFT 来计算频谱分量，然后在频谱分量上直接计算频谱相关性。用于频谱检测的循环稳态检测器的数字实现如图 7-3 所示。循环稳态检测的主要好处是它可以区分噪声能量和信号能量。这是因为噪声没有谱相关，而调制信号通常是循环稳态的，因为有嵌入的信号周期的冗余信息。较能量检测而言，循环稳态检测器应对噪声的不确定性更为稳健。而且，它可以在比能量检测更低的 SNR 条件下工作，因为后者无法利用接收到的信号里嵌入的信息，而特征检测器可以做到。众所周知循环稳态检测比能量检测可以获得更大的处理增益，可是循环稳态检测的实现更为复杂。因此，短时间的频谱空洞也许无法有效利用。当我们考虑多系统共存时，循环稳态检测器的系统辨识让系统更加复杂，所以我们需要一个有效的方法来处理。

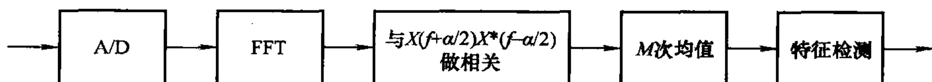


图 7-3　循环稳态检测器的数字实现

7.1.1.4　小波检测

在检测宽带信号时，与传统的多窄带带通滤波（BPF）方法比，小波方法在实现代价和灵活性两方面都有优势。为了辨识出空闲频段的位置，整个宽带被考虑成一串频率子带，其中在各个子带上功率谱特性是光滑的，只有在子带的边界上才会发生突变。通过对观察到的信号 $x[n]$ 功率谱密度（PSD）进行小波变换，功率谱密度 $S(f)$ 的奇点就可以被找出来，这样空闲的频段就找到了。在实际使用中，小波检测的一个严重挑战是大带宽带来的高采样率。小波检测器的数字实现如图 7-4 所示。当频谱内有多个系统共存时，小波检测很难分别出多个系统，因为小波检测器是用来辨识空闲频谱的边界的，它很难应付系统间的互相干扰。

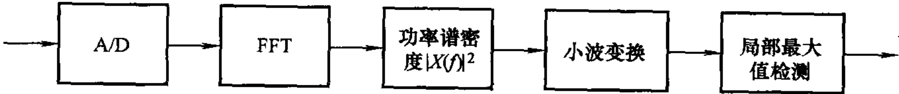


图 7-4 小波检测器的数字实现

表 7-1 中总结了上述频谱感知技术的优缺点。

表 7-1 频谱感知技术的优缺点

频谱感知技术	优　点	缺　点
能量检测	1. 不需要先验信息 2. 计算量小	1. 不能在低 SNR 下工作 2. 不能区分使用同一信道的用户
匹配滤波器	1. 最优检测性能 2. 计算量小	1. 需要主用户的先验信息 2. 需要对各种主系统信号分别设计
周期稳态检测	1. 低 SNR 下稳定 2. 对干扰稳定	1. 需要主用户的部分信息 2. 计算量大
小波检测	1. 有效检测宽带信号	1. 不能处理扩频信号 2. 计算量大

7.1.2 功率控制

当认知无线电在认知无线电网络里以次要用户身份运行时，它的频谱接入优先级比已经存在的主系统低。因此，认知无线电应该有防止主系统在传输中遭到有害干扰的设计。可是，严格保证主系统的运行不是容易的事，因为有隐藏终端问题，这在载波监听网络里已经充分研究了。当两个终端在一个基站范围内，可是互相无法监听，或者被某些对无线电信号透明的物理障碍分开，它们就被称为互相隐藏。在无线局域网里或类似的无线数据网络里，隐藏终端问题会在接收机造成可能的碰撞，当发射机是隐藏时（即无法感知到彼此的发射）。这可以通过四轮握手方法来解决，其中先发射一个短的请求信号，数据包只有在接收机同意的情况下才能发送。因此，碰撞只会在传输请求信号时发生，而不会在数据信号时发生。通过这个检验和观测的过程，隐藏终端问题得到了解决，正如 IEEE 802.11 MAC 里一样。

类似地，隐藏终端问题也会在认知无线电网络里发生。当认知无线电发射时，它需要知道它的发射对于主系统的影响。更具体地说，主接收机（PS-Rx）处的干扰应该保证在一定门限以下。我们把主系统接收机看成是一个隐藏终端，当其存在无法被认知无线电检测时。图 7-5 所示为认知无线电网络里分别因为路径损失和阴影效应的隐藏终端问题。认知无线电网络也许会因为隐藏终端问题而遭受严重损失，因为它将无法预知的干扰引入了主系统。

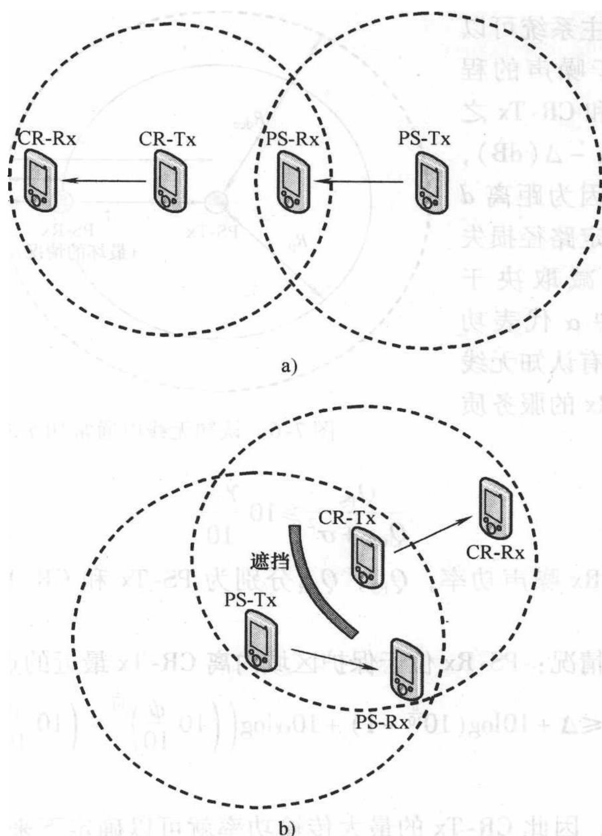


图7-5　隐藏终端问题的原因
a）路径损失　b）阴影效应

7.1.3　功率缩放控制

传统的频谱检测方法主要是被动地通过检测主系统发射机（PS-Tx）的信号。既然我们经常假设认知无线电发射机（CR-Tx）并不知道它到PS-Rx的信道，认知无线电的敏感度就应该远高于PS-Rx。一般来说，我们需要考虑最坏的情况，其中PS-Rx位于主系统覆盖范围内离CR-Tx最近的地方，如图7-6所示。相应的CR-Tx的传输功率控制，称为功率缩放控制，可以通过感知PS-Tx的信号来完成。

具体地说，我们假设主系统有最小的SINR需求γ_{dec}来在目标速率下成功译码。当没有干扰时，γ_{dec}出现在离PS-Tx半径为R_{dec}的范围内。这个想法是保证一定保护半径R_p内的PS-Rx，使得PS-Rx有确定的成功接收速率，甚至在有次要用户传输时。定义Δ(dB)为PS-Tx和半径R_{dec}间的信号衰落，μ(dB)为保护

余量，这代表着主系统可以容忍的干扰高于噪声的程度，d 为 PS-Tx 和 CR-Tx 之间的距离，$\psi = \lambda - \Delta$(dB)，其中 λ(dB) 为因为距离 d 的路径损失。假定路径损失和传播功率衰减取决于 $g(r) = r^{-\alpha}$，其中 α 代表功率损失指数。当有认知无线电传输时，RS-Rx 的服务质量可以保证为

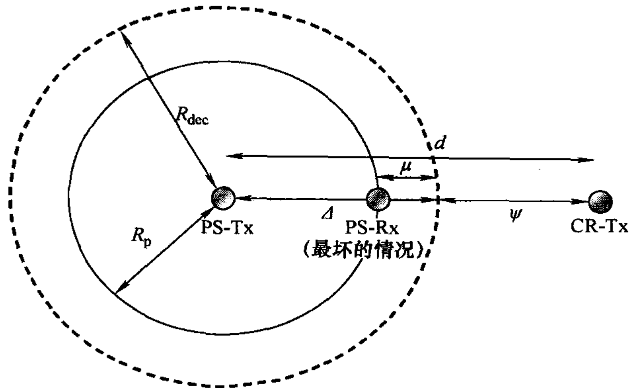


图 7-6 认知无线电通常功率缩放控制

$$\frac{Q_{PS}}{Q_{CR}+\sigma^2} \geqslant 10\frac{\gamma_{dec}}{10} \tag{7-8}$$

式中，σ^2 为 PS-Rx 噪声功率，Q_{PS}，Q_{CR}分别为 PS-Tx 和 CR-Tx 接收到的信号功率。

考虑最坏的情况：PS-Rx 位于保护区域内离 CR-Tx 最近的点，我们有

$$10\log\left(\frac{P_{CR}}{\sigma^2}\right) \leqslant \Delta + 10\log(10^{\frac{u}{10}} - 1) + 10\alpha\log\left(\left(10\frac{\psi}{10}\right)^{\frac{u}{10}} - \left(10\frac{u}{10}\right)^{\frac{1}{\alpha}}\right) \equiv f(\psi) \tag{7-9}$$

α 和 μ 均为常数，因此 CR-Tx 的最大传输功率就可以确定下来。请注意认知无线电的先验信息包括 σ^2、Δ、μ 和 α。所以，ψ 为频谱感知的目标，这可以通过估计的虚警概率获得。当考虑 CR-Tx 可能相对主信号被遮蔽时（信号损失为 β(dB)），CR-Tx 必须测量余量 $\mu+\beta$。调整方程以考虑阴影效应，我们有

$$10\log\left(\frac{P_{CR}}{\sigma^2}\right) \leqslant \Delta + 10\log(10^{\frac{u}{10}} - 1) + 10\alpha\log\left(\left(10\frac{\psi-\beta}{10}\right)^{\frac{1}{\alpha}} - (10^{\frac{u}{10}})^{\frac{1}{\alpha}}\right) \tag{7-10}$$

7.1.4 协作频谱感知

认知无线电对主系统频谱感知的主要挑战是隐藏终端问题，与任何传输感知机制一样，例如无线网络里常用的 CSMA。图 7-7 描述了频谱感知的隐藏终端问题。主系统的传输功率范围为左边的大圆，而右边的小圆代表认知无线电的传输范围。认知无线电发射机（CR-Tx）希望能够感知到频谱空洞，动态地接入信道进行传输，并且保证以低于 γ 的概率干扰主系统（理想情况下 $\gamma \to 0$）。可是，遮蔽衰落的某种阻碍结果阻止了认知无线电发射机的有效频谱感知，类似 CSMA 中的隐藏终端问题。立刻能想到的解决方法是采用协作频谱感知，即将一组传感器放到不同地点，来检测主系统发射机可能的传输，并将这样的检测信息从分布的

传感器中继到认知无线电发射机。这个开创性工作是由佐治亚理工学院首先进行的[8,9]。以下我们将把这个问题建模为传感器检测问题。

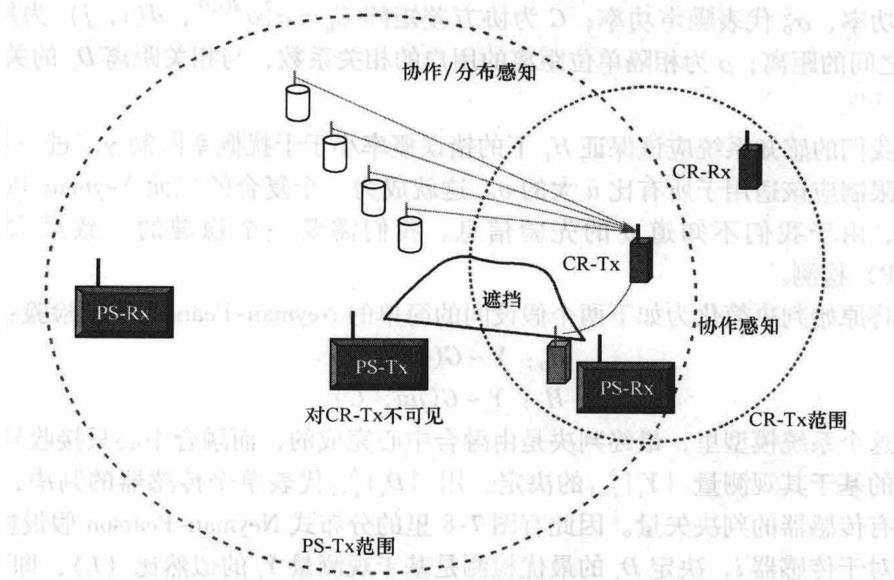


图7-7 频谱感知的隐藏终端问题和协作策略

考虑认知无线电系统中的用户协作，研究者们已经考虑了两种方案：①在所有协作用户中进行某种联合检测；②最终判决是基于每个协作用户的硬判决结果。根据文献［3］的方法，我们集中研究更可行的系统，其中单个次要用户在它们所检测的频带内作出独立的关于主信号是否存在的判决。它们将其判决发送给一个融合中心，基于这些判决作出最后判决。实际情况里，这个融合中心可以是一个中央控制器，为次要认知无线电节点/用户管理信道的分配和调度。这样的系统也可以帮助次要用户来交换它们的判决，而每一个用户自己进行所有判决的融合。

假定融合中心知道所有协作次要用户的地理位置，所以可以构造出它们观测量的相关性。可是主系统的位置是不知道的。感知问题现在就成了一个二元假设检验问题，来判决在接收机处平均接收到的功率是否高于传输范围边缘的预期功率。当主系统在传输时，次要用户在其范围内，它们接收到的功率是主系统信号功率与噪声功率之和。在这个情况下，我们假设接收信号功率与阴影衰落一样满足对数正态分布。

从 N 个协作传感器观测到的功率水平形成了观测矢量 Y。二元假设问题为

$$H_0:\ Y \sim G(0,\ \sigma_0^2 I)$$
$$H_1:\ Y \sim G(\theta u,\ C)$$

其中，$G(m, V)$ 为矢量高斯分布，其平均矢量为 m，协方差矩阵为 V；u 为单位矢量，I 为单位矩阵；θ 为 H_1 时的观测量均值，$u=\mathrm{E}[10\log_{10}(1+SNR)]$ 为平均总功率，σ_0^2 代表噪声功率；C 为协方差矩阵 $C_{ij}=\sigma_1^2\sigma^{d(i,j)}$，$d(i, j)$ 为用户 i 和 j 之间的距离；ρ 为相隔单位距离的用户的相关系数，与相关距离 D_c 的关系为 $\rho=\mathrm{e}^{-1/D_c}$。

我们的感知系统应该保证 H_1 下的错误概率小于干扰概率限制 γ。进一步地，这个限制应该适用于所有比 u 大的 θ。这就成为一个复合的二元 Neyman-Pearson 检测，由于我们不知道 θ 的先验信息。我们需要一个稳健的一致最大功效（UMP）检测。

将原始判决简化为如下两个假设间的简单的 Neyman-Pearson 假设检验：

$$H_0:\ Y\sim G(0,\ \sigma_0^2 I)$$

$$H_1:\ Y\sim G(\mu u,\ C)$$

这个系统模型里，最终判决是由融合中心完成的，而融合中心只接收到了传感器的基于其观测量 $\{Y_i\}_{i=1}^N$ 的决定。用 $\{D_i\}_{i=1}^N$ 代表单个传感器的判决，D 代表所有传感器的判决矢量。因此有图 7-8 里的分布式 Neyman-Pearson 假设检验。

对于传感器 i，决定 D_i 的最优检测是基于观测量 Y_i 的似然比（L），即

$$D_i=1_{\log[L(Y_i)]>\eta} \tag{7-11}$$

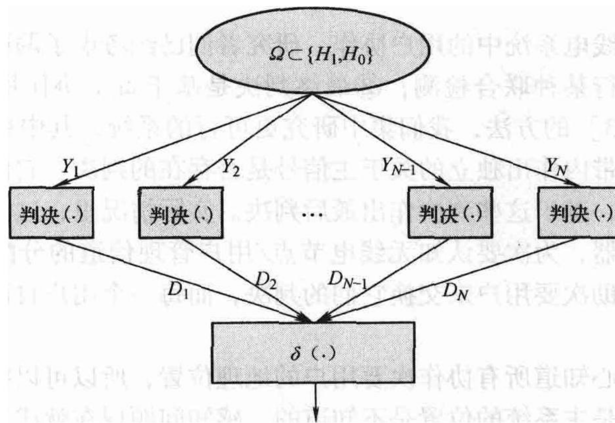


图 7-8 传感器检测中的协作感知

对于典型的传感器方法，每个传感器将其观测量化（决策），然后在融合中心作出最终判决 δ。对于这样的数据融合问题，最简单的次优解决方法之一就是计数规则（也称为选择规则），即计算出选择 H_1 的节点数量，然后和一个阈值比较。这样的方法更像协作中的译码转发。更多的次优方法可以在参考文献[5]里找到。

另外一个协作式频谱感知的方法如图7-9所示，对观测量采用自适应滤波来做出判决，这更像协作中的放大转发。

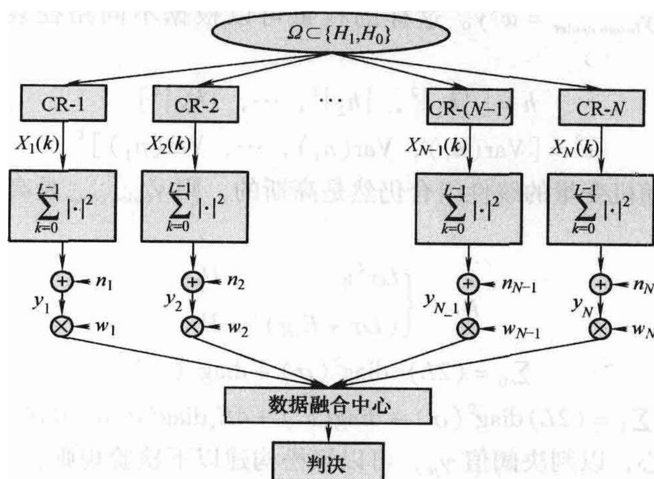


图7-9　数据融合中的协作感知

对于认知无线电网络中的这些协作式次要节点，在第 k 个时刻的频谱感知的二元假设检验为

$$H_0:\ x_i(k)=v_i(k),\ i=1,\ 2,\ \cdots,\ N$$
$$H_1=x_i(k)=h_is(k)i+v_i(k),\ i=1,\ 2,\ \cdots,\ N$$

式中，$s(k)$ 代表主用户传输的信号，h_i 为信道增益，$x_i(k)$ 为第 i 个协作式次要用户接收到的信号。信道里的噪声为加性高斯白噪声，因此 $v_i(k)\sim G(0,\ \sigma_i^2)$ 和 $s(k)$ 独立，其中 $\sigma=[\sigma_1^2,\ \sigma_2^2,\ \cdots,\ \sigma_N^2]^{\mathrm{T}}$。

每一个协作式次要用户都在 L 个采样的区间上进行非相关检测（L 由时间-带宽乘积决定）：

$$u_i=\sum_{k=0}^{L-1}|x_i(k)|^2\quad i=1,2,\cdots,N$$

以上式子表示在第 i 个协作式次要节点/用户上基于能量检测的局部感知，即高斯随机变量的二次方和。请回顾中心卡方分布，我们有

$$\frac{u_i}{\sigma_i^2}\sim\begin{cases}\chi_L^2, & H_0\\ \chi_L^2(\eta_i), & H_1\end{cases}$$

局部的信噪比定义为

$$\eta_i=\frac{E_s|h_i|^2}{\sigma_i^2}$$

一旦对每个节点定义了判决阈值，我们就很容易得到单个协作式认知无线

电的判决规则。可是，希望调节权重系数 $w=[w_1, w_2, \cdots, w_N]^{\mathrm{T}}$，使得可以基于观测值 $y=u+n$ 在融合中心达到最优频谱感知，然后对主系统的出现做出最终判决，即 $y_{\text{fusion center}}=w^{\mathrm{T}}y$。这样的权重可以根据不同路径衰落进行调制。我们定义

$$h=[\,|h_1|^2,\ |h_2|^2,\ \cdots,\ |h_N|^2]$$

$$v^2=[\mathrm{Var}(n_1),\ \mathrm{Var}(n_1),\ \cdots,\ \mathrm{Var}(n_1)]^{\mathrm{T}}$$

既然高斯随机变量的线性组合仍然是高斯的，则 $y_{\text{fusion center}}$是高斯的，其均值和方差满足

$$\mu_{fc}=\begin{cases} L\sigma^{\mathrm{T}}w & H_0 \\ (L\sigma+E_s g)^{\mathrm{T}}w & H_1 \end{cases}$$

$$\Sigma_0=(2L)\ \mathrm{diag}^2(\sigma)+\mathrm{diag}\ (v^2)$$

$$\Sigma_1=(2L)\mathrm{diag}^2(\sigma)+\mathrm{diag}(v^2)+4E_s\mathrm{diag}(g)\mathrm{diag}(\sigma)$$

在融合中心，以判决阈值 γ_{fc}，可以轻松构建以下检验规则：

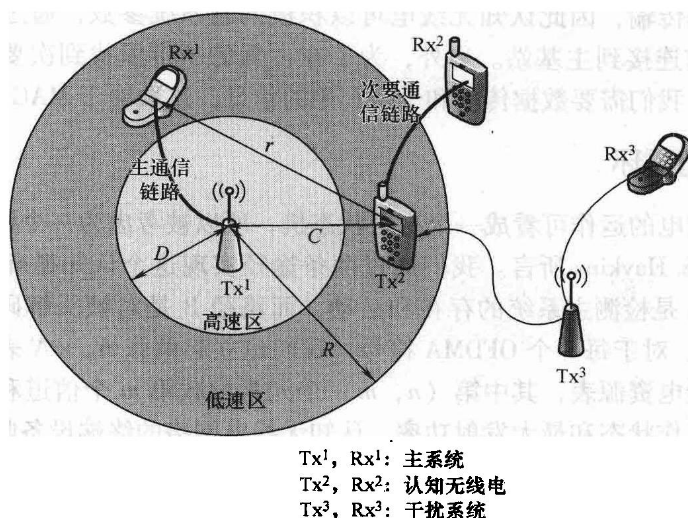
$$y_{\text{fusion center}} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma_{fc}$$

我们很容易计算其判决概率和虚警概率，作为检验的性能指标，这样就可以进行性能优化了。然而，请注意这里的一些信息在实际系统里可能难以获取，例如信道增益，尽管我们已经只要求了很少的信息来计算融合中心的均值和方差。更多的细节请参见参考文献［6］。事实上，在很多实际的共存多无线电系统中，我们不得不考虑频谱感知和其他认知无线电操作里的累积干扰。Ghasemi 和 Sousa 在参考文献［7］中已经总结了累积干扰的一个数学模型，并验证了其在协作式频谱感知里的应用。

7.2 认知 OFDMA 系统中的频谱感知

在介绍基本的频谱感知技术后，我们准备为实际的 OFDMA 系统设计频谱感知，利用参考文献［10］第 5 章的叠加速率-距离。考虑一个蜂窝结构的 OFDMA 系统。基站位于小区的中间，如图 7-10 所示，其覆盖半径为 R。令 W 为总的带宽，包含 N_{F} 个子带。每一个小区分配一个子带，并进一步划分为 N 个信道。类似地，时间轴被划分为 OFDMA 符号周期。然后，无线电资源，包括信道和 OFDMA 符号，被分配给小区内的活跃用户。这个信息通常在帧头里标明。不失一般性，我们以 IEEE 802. 16 中的帧结构为例子。每一帧以一个前导为开始，主要用于同步和信道估计，接下来是帧头，包括 DL_MAP 和 UL_MAP。另外，帧

头也定义了传输参数，例如前向纠错编码速率，下行数据块和上行数据块的调制方式。为了简化这个模型但不失一般性，假定有两个数据传输速率，遵循一定的自适应调制和编码（AMC）。虽然数据传输速率可以自适应地进行调节，基本的符号速率在绝大多数系统里是非常固定的。



Tx1，Rx1：主系统
Tx2，Rx2：认知无线电
Tx3，Rx3：干扰系统

图7-10　OFDMA 认知无线电系统模型

可以将认知无线电在主系统的一个子带里接收到的信号分为以下六类：

C_0：$w_{LP}(t)$ 只有背景噪声；

C_1：$S_t(t)$ 主系统数据信号；

C_2：$S_c(t)$ 主系统控制信号；

C_3：$i_s(t)$ 与主系统基本符号速率相同的干扰信号；

C_4：$i(t)$ 与主系统基本符号速率不同的干扰信号；

C_5：$i_c(t)$ 干扰的控制信号。

其中，$w_{LP}(t)$ 代表等价的基带（BB）加性高斯白噪声（AWGN），均值为零，双边功率谱密度为 N_0，$s_t(t)$ 代表主系统传输信号。另外，$i_s(t)/i(t)$ 代表干扰系统的传输信号，具有或没有和主系统相同的符号速率，而且假设在一个子带里是白色的。根据中心极限定理，传输信号可以建模为独立循环对称复高斯过程。同时，$S_c(t)$、$i_c(t)$ 代表控制和管理信号，分别来自主系统和干扰系统，并周期性发射。以下，我们将控制与管理信号简写为控制信号。通过分析接收到的信号，认知无线电可以判决主系统的工作状态，其中包括：

S_0（不存在）：C_1 和 C_2 都不存在

S_1（存在，但是不工作）：C_2 存在，C_1 不存在

S_2（存在，工作，高速率）：C_1 存在，数据在一个信道里以高速率传输

S_3（存在，工作，低速率）：C_1 存在，数据在一个信道里以低速率传输

S_4（存在，工作，待机）：C_1 存在，信道里没有数据

这里，我们进一步考虑状态 S_1（存在，但不工作），其中主网络用户存在，但是没有数据传输，因此认知无线电可以积极调制系统参数，通过协作式（多无线电）通信连接到主基站。另外，为了在占据的子带里找到次要用户可用的无线电资源，我们需要数据传输和子带利用的信息，这取决于 MAC 协议。

7.2.1 认知循环

认知无线电的运作可看成一个有限状态机，所以被考虑为一个认知循环，如 J. Mitola Ⅲ和 S. Haykins 所言。我们通过两条途径实现这个认知循环，如图 7-11 所示。路径 A 是检测主系统的存在和活动，而路径 B 是对帧头解码，并决定频率-时间利用。对于每一个 OFDMA 符号，我们建立起两张 $N_F \times N$ 表，包括信道状态表、无线电资源表，其中第（n，m）个元素记载第 m 个信道和第 n 个子带上主系统的工作状态和最大发射功率。认知无线电网络的终端设备收集感知到的信息，并做出工作方式的最优决策，包括 MAC 和 SDR 软硬件的设置、网络路由和切换、高层应用需求的维护。基于这样的系统体系，认知无线电网络可在未来的无线通信里得到应用。

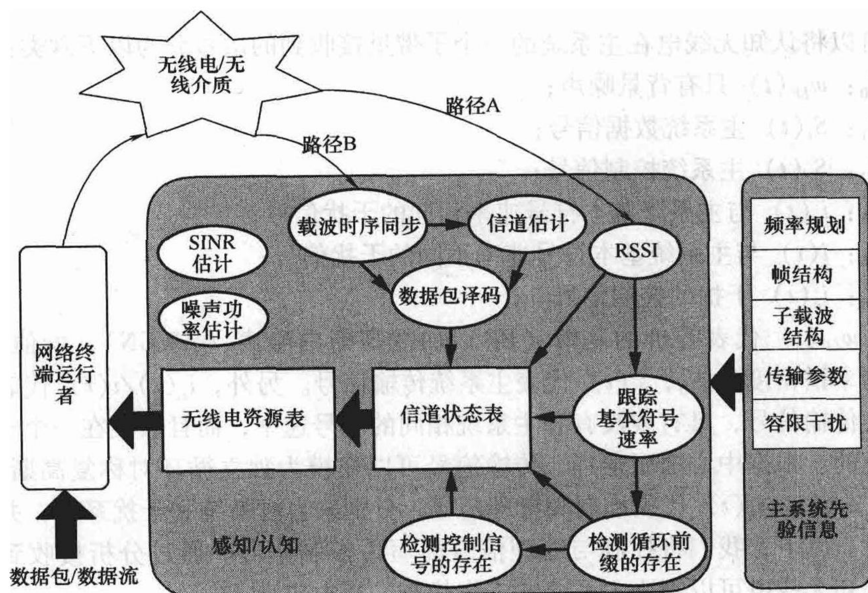


图 7-11 认知循环

7.2.1.1 认知循环——路径A

接收信号强度指标（RSSI）是一个在频谱感知里简单但广泛使用的方法。可是它不足以区分主系统的信号（C_1）和干扰信号（C_3和C_4）。例如在2.4GHz ISM频段里，大的RSSI可能是频带外的微波信号造成的，而不是IEEE 802.11b网络的信号。另外，噪声的不确定性也会严重降低能量检测的性能。因此，我们需要通过变换获得数据信号的特征来抑制噪声和干扰信号。考虑到基本符号速率不会变化或者变化是有限的这一事实（例如可升级的WiMAX），我们将数据信号建模为一个周期稳态信号。尽管初步的频谱感知是存在的，我们采用频谱线发生器来满足认知无线电的快速的需求。然而，在我们的系统模型中，C_1和C_3有相同的基本符号速率。通过OFDMA系统中的循环前缀（CP），我们可以准确地检测活跃主系统的存在。最后，既然基站周期性广播控制信号（例如IEEE 802.11中的信标信号），来维持移动网络的运作，我们可以通过这一性质来判定不活跃主系统的存在。

7.2.1.2 认知循环——路径B

在路径A中，我们判定主系统的状态：不存在（S_0），存在但不工作（S_1），存在并工作（S_2，S_3，S_4）。进一步考虑被传输所占据的信道，应用速率-距离的实质，使得同时通信变为可能。信道占据的信息和数据传输速率可以在认知无线电和主基站实现同步后，通过分析帧头获得。

下面将OFDMA系统频谱感知中的认知信息总结如下：

射频信号处理：（载波）频率，带宽，RSSI，信噪干扰比（SINR），噪声能量；

基带预检测信号处理：基本符号速率，载波和时刻，导频信号，信道衰减；

基带后检测信号处理：系统辨识，调制参数，纠错编码种类和速率；

网络处理信息：多址协议，无线电资源分配。

假定认知无线电知道可能的主系统的先验信息，包括频率规划、帧结构、子载波结构（FFT的规模，导频的位置）、传输参数（基本符号速率，循环前缀长度，纠错编码种类）和主系统的容限干扰，因为频谱使用是接受管理的，而系统规范都是充分定义了的。

7.2.2 主系统状态的区分

下一个设计步骤是在一个子带里通过RSSI、基本符号速率、循环前缀、控制信号，来判断主OFMDA系统的存在和活动，其中RSSI只是简单的能量检测器。一般情况下，我们不知道主系统存在的先验信息。所以，我们用Neymann-Pearson准则来设计频谱感知过程。

7.2.2.1 基本符号速率

如果存在主系统的数据信号，认知无线电接收到的信号可以写成

$$z(t)=\Re\left\{e^{j2\pi f_c t}\left(\sum_{n=-\infty}^{\infty}x(n)h(t-nT_s)+W_{LP}(t)\right)\right\}=$$
$$\Re\{e^{j2\pi f_c t}(s_t(t)+W_{LP}(t))\}=\Re\{y(t)\} \tag{7-12}$$

这里，f_c 表示载波频率，T_s 表示基本符号速率的倒数，$h(t)$ 表示脉冲整形滤波器，$x(n)$ 表示时域内传输的数据。或者

$$x[n+(N_{\text{FFT}}+N_{\text{CP}})m]=$$
$$\begin{cases}\dfrac{1}{\sqrt{N_{\text{FFT}}}}\displaystyle\sum_{k=0}^{N_{\text{FFT}-1}}X_m(k)e^{j2\pi}\dfrac{kn}{N_{\text{FFT}}} & n\in\{0,1,\cdots,N_{\text{FFT}}-1\}\\ x[n+(N_{\text{FFT}}+N_{\text{CP}})m+N_{\text{FFT}}] & n\in\{-N_{\text{CP}},-N_{\text{CP}}+1,\cdots,-1\}\end{cases}\quad(m\in\mathbb{Z}) \tag{7-13}$$

在式（7-13）里，$X_m(k)$ 表示调制后在第 m 个 OFDMA 符号的第 k 个子载波上的数据，N_{FFT}为 FFT 规模，N_{CP}为循环前缀的长度。不失一般性，$h(t)$ 假设为二次方根升余弦滤波，其滚降系数小于 100%，我们有

$$E[z^2(t)]\approx\frac{1}{2}\left(\frac{\sigma_x^2}{T_s}+\sigma_w^2\right)+\frac{\sigma_x^2}{T_s}\Re\{Z_1 e^{j2\pi\frac{t}{T_s}}\} \tag{7-14}$$

其中

$$Z_m=\frac{1}{2\pi}\int_{-\infty}^{\infty}H(jw)H^*(j(w-2\pi m/T_s))\,dw \tag{7-15}$$

噪声功率为 σ_w^2，$Z_{-m}=Z_m^*$ 和 $Z_0=1$。注意基本符号速率的元素与噪声功率无关，所以能量检测器里噪声不确定性的问题就不存在了。可是，在式（7-14）中的第二项内接收到的信号功率减少了 Z_1。所以，检测器的性能和 Z_1 相关。从式（7-15）来看，我们得出这样的结论：如果 $h(t)$ 有更大的带宽（即更高的滚降系数），检测器就有更好的性能。类似的现象也发生在时间跟踪系统内。进一步地，既然这个算法可以在 A/D 转换器前的射频（RF）上实现，它可以做得更快。

7.2.2.2 OFDMA 信号的循环前缀

检测循环前缀的初始设置类似于参考文献［38］所述。将 $2N_{\text{FFT}}+N_{\text{CP}}$个采样（采样率 $1/T_s$）收集起来。假定这个区域包含一个完全的 OFDMA 符号。检测问题就变成了

$$H_0:\ r(n)=w_0(n)\quad n=0,\ 1,\ \cdots,\ 2N_{\text{FFT}}+N_{\text{CP}}-1$$
$$H_1:\ r(n)=s(n)+w_1(n)\quad n=0,\ 1,\ \cdots,\ 2N_{\text{FFT}}+N_{\text{CP}}-1$$

在 H_1 下因为没有同步，存在时间漂移 ϑ 和频谱漂移 ε。于是，主系统的传输信号就成了

$$s(n)=x(n-N_{\text{CP}}-\theta)e^{j2\pi\varepsilon n/N_{\text{FFT}}}$$

令 I 和 $\tilde{I}$ 为两个采样区间，包括循环前缀和它的复制，即

$$I=\{\theta,\ \theta+1,\ \cdots,\ \theta+N_{\mathrm{CP}}-1\}$$

$$\tilde{I}=\{\theta+N_{\mathrm{FFT}},\ \theta+N_{\mathrm{FFT}}+1,\ \cdots,\ \theta+N_B-1\}$$

如果主系统存在，循环前缀里的采样和它们的复制是相关的，或者

$$E[r(n)r^*(n+m)\mid H_1]=\begin{cases}\sigma_s^2+\sigma_w^2 & m=0\\ \sigma_s^2\mathrm{e}^{-\mathrm{j}2\pi\varepsilon} & m=N_{\mathrm{FFT}},\ n\in I\\ 0 & \text{其他}\end{cases}$$

将 $w_1(n)$ 建模为白高斯过程，其均值为0，方差为 σ_w^2，而且 $\mathrm{E}[\,|s(n)|^2]=E[\,|x(n)|^2]=\sigma_s^2$。另一方面，$w_0(n)$ 代表干扰信号以及噪声的叠加，被建模为白高斯过程，其均值为0，方差为 $\sigma^2=\sigma_s^2+\sigma_w^2$，即

$$E[r(n)r^*(n+m)\mid H_0]=\begin{cases}\sigma^2 & m=0\\ 0 & \text{其他}\end{cases}$$

假定在两个假设下总的功率都是一样的，这样就导致最坏的情况，因为这时能量检测已经无法进行。然后，似然比检测（LRT）变成

$$M(r)=|S(\theta)|\cos(2\pi\varepsilon+\angle S(\theta))-\frac{\rho}{2}P(\theta)\ \underset{H_0}{\overset{H_1}{\gtrless}}\ \tau_{\mathrm{CP}} \tag{7-16}$$

其中

$$S(\theta)=\sum_{n\in I}r(n)r^*(n+N_{\mathrm{FFT}})$$

$$P(\theta)=\sum_{n\in I}[\,|r(n)|^2+|r(n+N_{\mathrm{FFT}})|^2]$$

$$\rho=SINR/(SINR+1)$$

以及 $SINR=\sigma_s^2/\sigma_w^2$。决策 $M(r)$ 只依赖于循环前缀中的采样和它们的复制，其能量 $P(\theta)$ 和相关 $S(\theta)$ 被纳入考虑。当参数（即 θ、ε、σ^2、$SINR$）未知时，这变成了复合检测问题，我们希望找到一致最大功效（UMP）检测。可是 UMP 检测并不存在，因为决策区域依赖于 θ。当没有 UMP 检测时，可以使用广义似然比检测（GLRT）。假设 $SINR$ 是已知的，而且有

$$M(r)=\frac{\sum_n |r(n)|^2}{\sum_n |r(n)|^2-\frac{2\rho}{1-\rho^2}\max_\theta\left\{|S(\theta)|-\frac{\rho}{2}P(\theta)\right\}}\ \underset{H_0}{\overset{H_1}{\gtrless}}\ \tau_{\mathrm{CP}} \tag{7-17}$$

注意，这个检测器只不过是两个假设下总能量的估计的比，而频率偏移并不影响式（7-17）。另外，我们计算虚警概率以达到

$$P_F=\int_{\tau_{\mathrm{CP}}}^{\infty}f(M(r)\mid H_0)\mathrm{d}M\approx Q\left(\frac{-u_1-u_2}{\sqrt{\sigma_1^2+\sigma_2^2}}\right) \tag{7-18}$$

式中，$f(\cdot)$ 为概率密度函数，$Q(x)$ 代表高斯随机变量（零均值，单位方差）的右尾概率，而且

$$\mu_1=(2N_{FFT}-N_{CP})(1-\tau_{CP})$$
$$\sigma_1^2=(2N_{FFT}-N_{CP})(1-\tau_{CP})^2$$
$$\mu_2=2N_{CP}-\frac{2N_{CP}-\rho\sqrt{\pi N_{CP}}}{1-\rho^2}\tau_{CP}$$
$$\sigma_2^2=2N_{CP}-\frac{2N_{CP}-4N_{CP}\sqrt{\pi}\left(\sqrt{N_{CP}+1}-\sqrt{N_{CP}}\right)}{1-\rho^2}\tau_{CP}+$$
$$\frac{2N_{CP}+4N_{CP}\rho^2\left(1-\frac{\pi}{4}\right)-4N_{CP}\sqrt{\pi}\left(\sqrt{N_{CP}+1}-\sqrt{N_{CP}}\right)}{(1-\rho^2)^2}\tau_{CP}^2 \qquad (7\text{-}19)$$

对 $Q(x)$ 取反，可以获得 NP 准则下的最优门限。下面，我们考虑特殊情况，$\rho=0$，即对白随机信号在未知噪声功率的高斯白噪声情况下的检测，检测统计量变成 $M(r)=1$。这就解释了为什么简单的能量检测在未知噪声功率下无法工作的原因。最后，为了处理未知 $SINR$，注意，式（7-18）和式（7-19）都是通过 ρ 与 $SINR$ 相关，所以，我们设计一个在目标 $SINR$、$\overline{SINR}$下的稳健的循环前缀检测器。参数 ρ，τ_{CP}可以实现通过 $SINR=\overline{SINR}$进行计算。进一步地，因为这个检测器利用目标 OFMDA 系统的循环前缀性质，所以它可以用来区分不同系统参数的 OFMDA 系统、N_{FFT}和 N_{CP}（例如 5GHz 无牌照频段的 IEEE 802.11a 和 IEEE 802.16）。

7.2.2.3 控制与管理信号

我们不指定可能的主系统的控制信号，但将其模型化为白高斯过程。因此，这个检测器对系统变化是稳健的。主系统的控制信号被建模为白高斯过程，其传输周期为 M_C 个采样，持续时间为 N_C 个采样。为了检测到控制信号，我们收集 M_C 个采样，并假设这个区间包括了一个控制信号持续时间。这样，检测问题就成了

$$H_0:\ r(n)=w(n) \qquad n=0,1,\cdots,M_C-1$$
$$H_1:\ r(n)=\begin{cases}w(n) & n\notin Q\\ S_c(n)+w(n) & n\in Q\end{cases}$$

式中，Q 表示采样周期，包括了控制信号，或者 $Q=\{\phi,\phi+1,\cdots,\phi+N_C-1\}$，其中 ϕ 代表控制信号的时间偏移。注意，频率偏移被忽略了，因为它只影响了相位偏移，而似然函数中只有接收到的信号幅度提供了信息。另外，在不同假设下接收到的信号为

$$E[r(n)r^*(n+m)|H_1]=\begin{cases}\sigma_s^2+\sigma_w^2 & m=0,\ n\in Q\\ \sigma_w^2 & m=0,\ n\notin Q\\ 0 & 其他\end{cases}$$

并且

$$E[r(n)r^*(n+m)\,|\,H_0] = \begin{cases}\sigma_w^2 & m=0\\ 0 & 其他\end{cases}$$

GLRT 变成

$$M(r) = \max_{\phi}\left\{\sum_{n=\phi}^{\phi+N_C-1} |r(n)|^2\right\} \underset{H_0}{\overset{H_1}{\gtrless}} \tau_C$$

这个检测器首先找到时间偏移，并测量在控制信号所在区间接收到的信号的能量，接着，我们放宽已知噪声功率的假设，这在检测主系统控制信号与干扰或者功率未知的背景噪声重叠时成立。除了噪声功率，信号功率是在 H_1 下另一个未知的参数。所以，我们有

$$M(r) = \max_{\phi}\left\{M_C\ln\left(\sum_{n}|r(n)|^2\right) - N_C\ln\left(\sum_{n\in Q}|r(n)|^2\right) - (M_C - N_C)\ln\left(\sum_{n\notin Q}|r(n)|^2\right)\right\} \underset{H_0}{\overset{H_1}{\gtrless}} \tau_C$$

最后，考虑一种非目标系统存在，而且广播控制信号的情况。如果认知无线电仅仅测量接收到的能量，它们将不能区分这种系统和主系统。但是，我们可以合理假设不同系统的控制信号的周期是不同的，因为针对特定的目的相应制定了不同的标准（例如，覆盖范围和数据率），这就导致了拥有各自的定时参数。因此，认知无线电能够通过追踪主系统控制信号的周期来检测控制信号。下面把控制信号重写如下：

$$r(n) = S_c(n)\sum_{m=-\infty}^{\infty} g(n - mM_C) + w(n) \quad n = 0,1,\cdots,LM_C - 1$$

其中，

$$g(n) = \begin{cases}1 & n=0,\ 1,\ \cdots,\ N_C-1\\ 0 & 其他\end{cases}$$

L 表示观察到的周期数。为了追踪控制信号的周期，我们采用一种类似基本符号速率追踪的方法。也就是，先将接收到的信号进行二次方，然后通过傅里叶变换从中提取控制信号基频 $1/M_C$ 的幅度信息。容易得到，

$$\mathrm{E}\left[\,|r(n)|^2\right] = \sigma_S^2\sum_{-\infty}^{\infty} g(n - mM_C) + \sigma_w^2\ n = 0,1,\cdots,LM_C - 1$$

我们注意到上面的这个函数是一个以 M_C 为周期的周期函数，其在频率 $1/M_C$ 的幅度是

$$L\frac{\sin(\pi N_C/M_C)}{\sin(\pi/MC)}\sigma_s^2$$

为了获得更高的吞吐量，控制信号传送得并不频繁，所以 M_C 很大，这就导致了很长的感知持续时间。不过，由于主基站被合理假设为固定的，这个过程仅仅在初始阶段使用。此外，这个检测器也能被用来检测在前缀里含有补零的基于 OFDM 的信号，例如多频带 OFDM。但是，在根据接收能量检测控制信号时，我们假设观测区间包含一个控制信号，这只在 $M_C >> N_C$ 时成立；对于基于 OFDM 的信号，$M_C = N_{FFT} + N_{CP}$，$N_C = N_{FFT}$，以上的条件不满足。要解决这个问题，我们要收集 $2N_{FFT} + N_{CP}$个采样，就像我们给前缀循环检测器做的那样，最优检测器有如参考文献［21］所述的类似结构。另一方面，我们也能通过追踪它们的符号周期来检测基于 OFDM 的信号，这与检测基本符号速率类似。

7.2.2.4 次要系统拥有的无线电资源

为了利用速率-距离特性，我们重新考虑图 7-10 中的距离关系和主通信链路使用的那些信道。这些信道是通过分析帧头获得的。$SINR_{min}$表示主接收机保持当前链路质量时的最小 $SINR$。注意，这个值取决于调制方法、FEC 类型和速率。因此，我们得到一个不等式

$$SINR_1 = \frac{G_{11}P_1}{G_{21}P_2 + N_1} \geqslant SINR_{min}$$

式中，P_i 表示在 T_X^i 的传输功率，N_i 表示在 R_X^i 的噪声功率，G_{ij}表示从 T_X^i 到 R_X^i 的能量损失。这样我们有

$$P_{2,max} = \frac{1}{G_{21}}\left(\frac{G_{11}P_1}{SINR_{min}} - N_1\right) = \frac{I_{CR}}{G_{21}} \quad (7\text{-}20)$$

式中，I_{CR}代表主系统在这个信道里能承受的最大干扰。由于 I_{CR}取决于主通信链路的信号质量，我们通过传输速率粗略地推算它，并认为它已知。另外，“距离”实际上是一种对接收到的信号功率的测量，我们必须明确接收到的信号功率和传播距离之间的关系。不失一般性，我们采用包含对数-距离路径损失和对数-正态阴影效应的大范围路径损失[28]，这么选择是因为频率感知不考虑瞬时信号接收。两个节点间的功率损失 G 如下：

$$G = Kd^{-\alpha}10^{\beta/10}$$

式中，K 是归一化常数，d 是两个节点间的距离，α 是路径损失指数，β 是阴影效应参数。阴影效应参数设计成一个均值为零、方差为 σ_β^2 的高斯随机变量。接下来，根据 T_X^2 和 R_X^1r 之间的距离信息是否可用来决定无线电资源，这对应上行链路和下行链路环境。

7.2.2.5 无线电资源——上行链路

在上行链路中，R_X^1 是主基站。认知无线电可以通过接收到的信号前缀的强

度来跟踪自己的位置。这样，我们得到如下的不等式

$$KP_2r^{-\alpha}10^{\beta/10}\leqslant I_{CR}$$

因为阴影效应因子β的缘故，干扰程度是一个随机变量，所以它只能在统计意义上得到保证。最佳准则是最大化传输功率使得与主系统的干扰概率小于ξ，也就是

$$P_{2,\max}=\arg\max_{P_2}\{Pr(KP_2r^{-\alpha}10^{\beta/10}>I_{CR})\leqslant\xi\}$$

这个准则和 Neyman-Pearson 准则拥有相同的理念，所以有

$$P_{2,\max}^{UL}=\frac{I_{CR}r^{\alpha}}{K}10^{-\sigma\beta Q^{-1}(\xi)/10}$$

注意，$P_{2,\max}^{UL}$和I_{CR}成比例。

7.2.2.6 无线电资源——下行链路

在下行链路中，R_X^1 成为主基站，因此这使得认知无线电很难估计主基站的位置。由于假设主基站在覆盖范围内均匀分布是合理的，利用贝叶斯方法，我们有

$$Pr\{P_2G_{21}>I_{CR}\}=\int_0^{\infty}Q\left(\frac{1}{\sigma_\beta}10\log_{10}\left(\frac{I_{CR}r^{\alpha}}{P_2K}\right)\right)\mathrm{d}F(r)$$

式中

$$F(r)=\begin{cases}\dfrac{A_H(r,c,D)}{\pi D^2} & 高速区\\[2ex] \dfrac{A_L(r,c,D,R)}{\pi(R^2-D^2)} & 低速区\end{cases}\tag{7-21}$$

$$A_H(r,c,D)=\begin{cases}\pi r^2U(D-c) & 0\leqslant r<|D-c|\\ \theta_2r^2+\theta_1D^2-cv & |D-c|\leqslant r<D+c\\ \pi D^2 & D+c\leqslant r\end{cases}$$

$$A_L(r,c,D,R)=A_H(r,c,R)-A_H(r,c,D)$$

$U(x)$ 是单位阶跃函数，

$$u=\frac{c^2+D^2-r^2}{2c}$$

$$v=\sqrt{D^2-u^2}$$

$$\theta_1=\cos^{-1}(u/D)$$

$$\theta_2=\cos^{-1}((c-u)/r)$$

其中，θ_1、$\theta_2\in[0,\ \pi]$，这样

$$P_{2,\max}^{DL}=\underset{P_2}{\arg}\left\{\int_0^{\infty}Q\left(\frac{1}{\sigma\beta}10\log_{10}\left(\frac{I_{CR}r^{\alpha}}{P_2K}\right)\right)\mathrm{d}F(r)=\xi\right\}$$

能够得到数值解。通过类似的步骤，我们能够容易地推广到多层数据传输速率。

现在，我们从先验信息得到了I_{CR}的粗略范围。不过，它能通过根据主系统的瞬

时链路质量动态计算。如果认知无线电能够和主系统交换信息，即协同共存，在和主基站同步以后，就能使无线电资源产生更高的效率，从而达到更大的信道容量。

7.2.3　频谱感知的步骤

频谱感知算法一个一个子带连续地感知感兴趣的整个频谱。为了快速地使用子带，我们首先要通过接收到的能量区分主系统的流量信号与噪声、干扰，并提取信号特征，信号特征包括基本符号速率和 CP。然后，确认活跃的主系统的存在。接着，我们通过三个条件下的控制信号考虑当前处于静默状态的主系统。这三个条件是：背景噪声（C_0）、干扰数据信号（C_3，C_4）和干扰控制信号（C_5）。当一个活跃的主系统存在时，我们获得沿着时间轴的每个信道的使用状态，以期望能通过译码帧头来进一步增强频谱效率，之后决定分配给次要系统的无线电资源。这个通用的感知算法在调整系统参数（T_s，N_{FFT}，N_{CP}，M_C，N_C）后可以应用到 IEEE 802.16 和其他 OFDMA 系统，总结如下：

1. 首先，设置信道状态表为 S_0，并重置无线电资源表。

2. 从 $n=1$ 到 N_F（子带），做如下步骤，然后转到步骤 3。

2.1　测量 RSSI，并区分假设检测

H_0：数据信号不存在

H_1：数据信号存在

如果 H_0 正确，跳转到步骤 2.6，否则转到步骤 2.2。

2.2　跟踪主系统基本符号速率

H_0：具有主系统基本符号速率的数据信号不存在

H_1：具有主系统基本符号速率的数据信号存在

如果 H_0 正确，跳转到步骤 2.7，否则转到步骤 2.3。

2.3　通过 OFDMA 信号的 CP，分离非协同干扰和主系统的数据信号

H_0：具有主系统 CP 特性的数据信号不存在

H_1：具有主系统 CP 特性的数据信号存在

如果 H_0 正确，跳转到步骤 2.7，否则转到步骤 2.4。

2.4　与主基站同步，包括载波和时序同步以及通过前导的信道估计。然后，跳转到步骤 2.5。

2.5　译码帧头（这里是 DL_MAP 和 UL_MAP），得到主系统的传输参数，包括 FEC 速率、调制方法和资源分配。这些参数用来设置信道状态表的第 n 行从 S_2 到 S_4。然后，转到步骤 2.9。

2.6　测量控制信号的能量，并检测假设检验

H_0：控制信号不存在

H_1：控制信号存在

如果 H_0 正确，跳转到步骤2.9，否则转到步骤2.8。

2.7　从非协同干扰中提取主系统的控制信号

H_0：主系统的控制信号不存在

H_1：主系统的控制信号存在

如果 H_1 正确，将信道状态表的第 n 行设置为 S_1。然后跳转到步骤2.9。

2.8　跟踪控制信号的周期，区分主系统的控制信号和其他系统信号

H_0：主系统的控制信号不存在

H_1：主系统的控制信号存在

如果 H_1 正确，将信道状态表的第 n 行设置为 S_1。然后跳转到步骤2.9。

2.9（可选）通过基本符号速率、OFDMA 系统的周期特性和控制信号鉴别系统。这些在认知和协作网络中是至关重要的信息。然后跳转到步骤2。

3. 根据信道状态表，设置无线电资源表。结束感知过程。

在步骤2.4和2.5中，如果由于缺乏帧信息，不能实现同步或包译码，这将导致性能受损的情况。在这当中，S_2、S_3 和 S_4 合并成一个状态，因此我们只需要两个 $N_F\times 1$ 表格。为了说明频谱感知步骤，我们建立一个感知树，如图7-12所示。这里，我们假设系统的控制信号和数据信号不能同时存在。这样，在认知无线电中接收到的信号存在15种可能的组合。

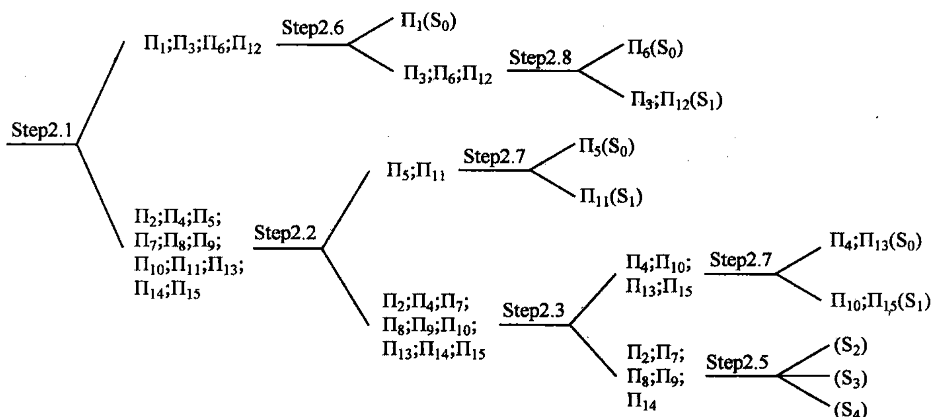


图7-12　感知树

$\Pi_1=\dot{C}_0$；

$\Pi_2=C_0\cap C_0$；$\Pi_3=C_2\cap C_0$；$\Pi_4=C_3\cap C_0$；$\Pi_5=C_4\cap C_0$；$\Pi_6=C_5\cap C_0$；

$\Pi_7=C_1\cap C_3\cap C_0$；$\Pi_8=C_1\cap C_4\cap C_0$；$\Pi_9=C_1\cap C_5\cap C_0$；$\Pi_{10}=C_2\cap C_3\cap C_0$；

$\Pi_{11}=C_2\cap C_4\cap C_0$；$\Pi_{12}=C_2\cap C_5\cap C_0$；$\Pi_{13}=C_3\cap C_4\cap C_0$；

$\Pi_{14}=C_1\cap C_3\cap C_4\cap C_0$；$\Pi_{15}=C_2\cap C_3\cap C_4\cap C_0$

如图所示，也许有几个可能的信号属于相同的状态。由于必须考虑其他可能系统的传输，一个类似的感知过程也可以被应用到其他的可能系统。之后再结合这些信道状态表和无线电资源，我们得到完整的频谱使用信息和相应最大传输功率。最后，完成这个感知/认知循环，图 7-13 是频谱感知的框图。在每个方框中标注了相应的功能。

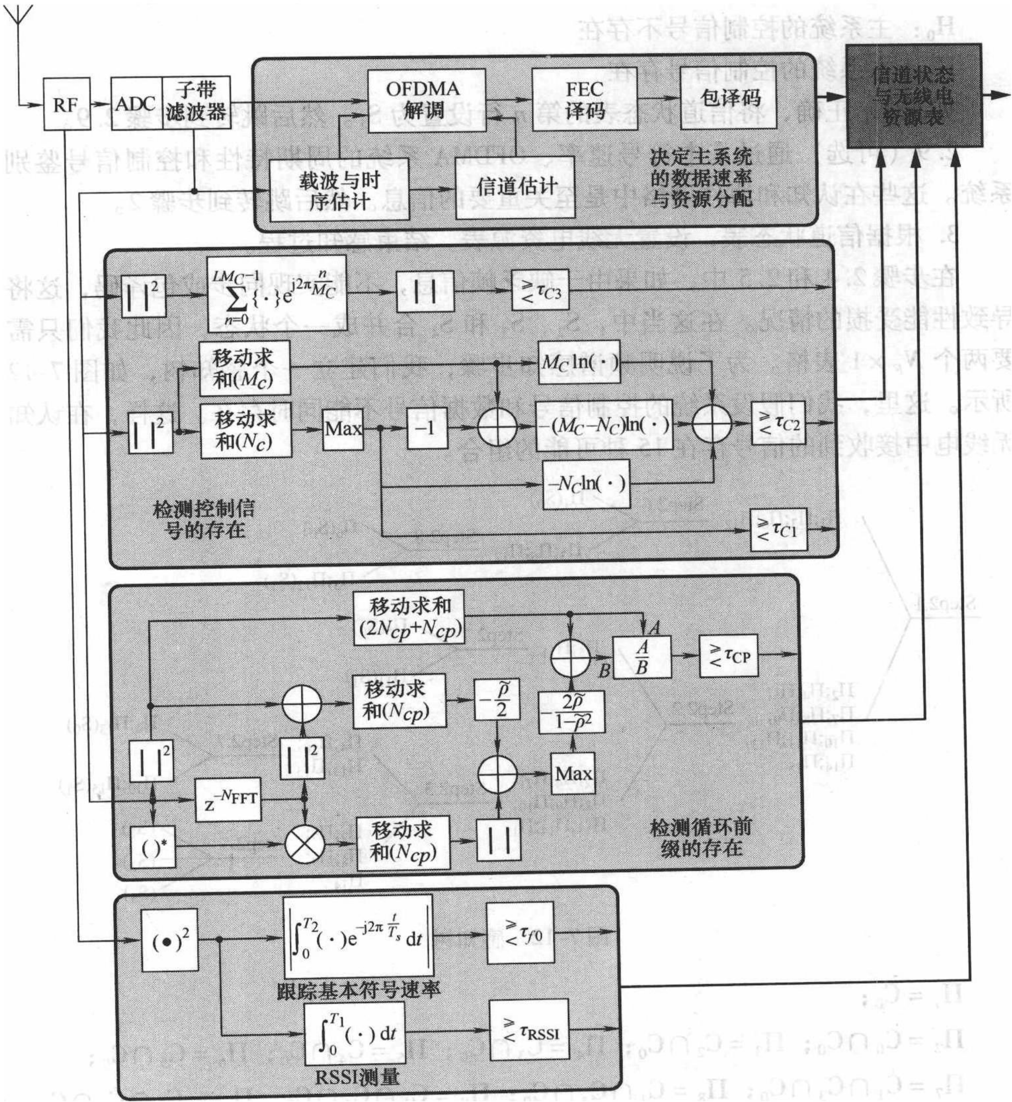


图 7-13　OFDMA 系统的频谱感知框图

通过利用速率-距离特性来达到认知无线电网络（超越传统的认知无线电链路）的有效频谱感知，我们建议为 OFDMA 系统获得一个认知信息的集合。为了避免对重叠的主系统形成干扰，认知无线电需要确定主系统的运行状态，包括频带、存在性、活跃性、资源分配和数据传输率。因此频谱感知已经被推广到多状态的辨别过程，图 7-12 中的感知树说明了这点。在讲述现存的能量检测和稳态特征检测时，我们设计了最优检测器去识别主系统在通用系统模型下的运行模式。仿真结果显示满足系统要求（$P_F<0.05$，$P_D>0.95$，$SINR=0\text{dB}$），这个结果是通过在加性高斯白噪声信道和频率选择性衰落信道中通过增加感知持续时间得到的。

在鉴别了主系统的运行状态后，我们推广到一般的认知无线电，并且通过考虑速率-距离本质来侦测在主系统忙碌期间的无线电资源，从而增加使用频带的机会。图 7-14 所示为次要系统所有的无线电资源关键取决于认知无线电可用的信息（例如主系统的链路质量和距离信息），同时它也和主系统能承受的干扰程度成比例。这大大地推广了以组网为目的的频谱感知的范围。

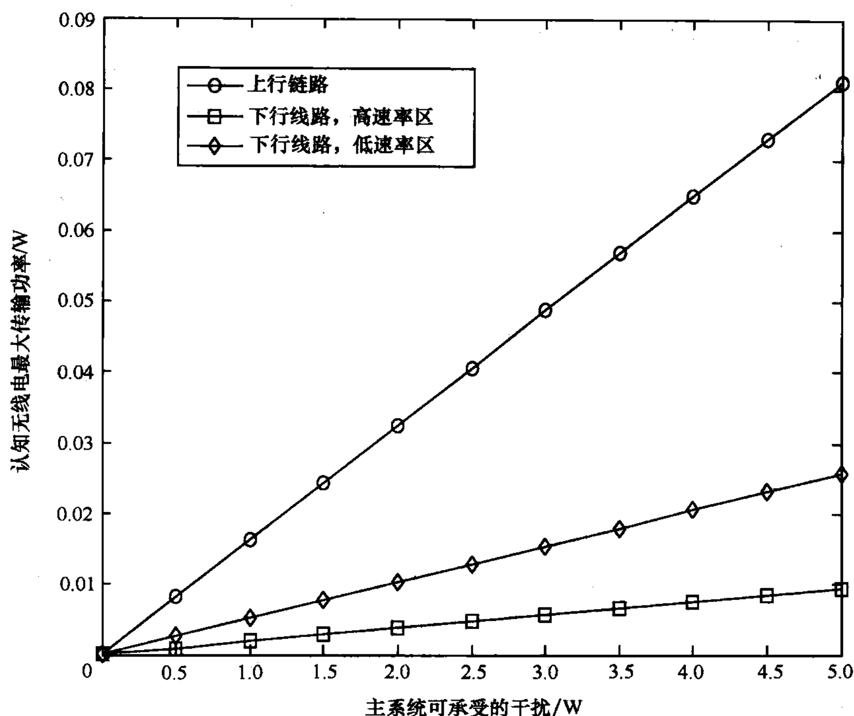


图 7-14 主系统的可承受干扰与次要系统的可用无线电资源

此外，这些方法能被运用于多无线电系统的鉴别。多无线电系统的鉴别是认知和协作网络至关重要的信息。通过调整认知无线电感知参数，我们之前提供的感知方法能推广到经典的 OFDMA 系统中，甚至推广到大部分移动通信系统中，当然这需要利用 MIMO 处理技术。

7.3 认知多无线电网络中的频谱感知

频谱感知技术传统上主要是为了单目标系统发展的，用来决定这个系统是否活跃或者发现没有被使用的频谱的位置（链路级别传输的频谱空洞）。另一方面，通常的功率缩放控制也是为单目标系统设计用来决定认知无线电的最大传输功率的。由于在认知无线电网络中，协作路由和网络效率优化都在多无线电环境中完成，认知无线电的感知也应该在多个共存无线电系统的基础上建立。在这样的多无线电的情形中，由于系统间的干扰，这些通常的频谱感知和功率控制技术是不够的。基于能量的方法只能区分通信系统是否是活跃的而不能识别出它们是哪个。因为对于各种目标系统需要不同的匹配滤波器，复杂性变成一个巨大的挑战。针对多系统稳态检测器的识别遇到的问题是我们需要从预先设定的库中高效地判定活跃的系统。小波检测器用来识别未用频谱的位置，因此用它来处理系统间的干扰比较困难。因此，一个更可靠的更通用的多系统感知算法才能应对这些挑战。我们在 7.3.1.2 节提供了一个通用的多系统感知算法。

在通常功率缩放控制方法中，认知无线电是被动地监听在目标系统中传输的信号。认知无线电的多预先存在系统的传输功率控制变得复杂，因为我们需要保证所有预先存在的系统都在可以接受的干扰下，特别是那些受累于隐藏终端问题的系统。我们将介绍一种积极感知算法来改善系统的容量，并在 7.3.2 节扩展到多系统。这将作为介绍包含多共存无线电系统的通用认知无线电感知的重要的一步[11]。

7.3.1 多系统感知

为了开发适用于多无线电的认知无线电网络的频谱感知，我们描述一种通用的多系统感知算法，这种算法利用特定系统的特点来识别多个活跃的系统。因为单独的能量检测不可靠，我们先通过周期性地利用脉冲整形滤波器来识别备选通信系统的基频。为了完成多系统感知，我们必须进一步利用特有的功率谱密度模式。如果添加的噪声是协方差矩阵未知的有色高斯噪声，我们可能利用更高阶的统计数据（四阶累积）来确保我们算法的成功。

7.3.1.1 问题表述

为了达到协作设置路由和优化这个网络的效率，认知无线电在认知无线电网

络中必须感知周围的通信系统/网络。我们称它为系统/网络感知。但是，因为系统间的干扰，通常的频谱感知技术对于 2.3 节中描述的多系统环境是不够的。因此，我们提供一个更可靠更通用的多系统感知算法。

假设有 Q 个备选通信系统，每个系统的传输信号都经过一个平坦的不相关的瑞利衰落信道，也就是每个信号有独立的复幅度 $a_i=|a_i|e^{j\theta_i}$，其中幅度 $|a_i|$ 是瑞利分布，$E\{|a_i|^2\}=\gamma_i^2$，相位 θ_i 在 $[0, 2\pi]$ 上均匀分布。此外，具有零均值，方差为 σ_w^2 的高斯白噪声 $w(t)$ 加到接收的信号。假设系统的活跃在每个感知的期间是不变化的。在活跃的系统数目为 P 的情况下（$P<Q$），接收到的无线电信号表述为

$$r(t)=Re\left\{\sum_{i=1}^{P}a_i s_i(t)+w(t)\right\}=Re\{y(t)+w(t)\} \tag{7-22}$$

式中，$S_i(t)$ 是第 i 个活跃系统的信号。

$$y(t)=\sum_{i=1}^{P}a_i s_i(t) \tag{7-23}$$

多系统感知的目标是判定处于活跃状态的系统的数目 P，并识别它们。

7.3.1.2 通用多系统感知算法

图 7-15 所示是通用多系统感知算法的框图。总结感知算法如下：

1. 能量检测和载波锁定来初始化算法。

2. 二次方收到的无线电信号，并通过一个包含所有基频的窄带带通滤波器过滤。检测基频，识别相应系统。

3. 估计目标系统功率谱密度。

4. 如果步骤 1 的结果是无，结束；否则，转到步骤 4。

5. 如果噪声的协方差矩阵已知，转到步骤 5；否则，转到步骤 6。

6. 将频谱估计结果单值分解并识别系统，结束。

7. 估计目标系统的三谱矩阵。将三谱矩阵 EVD，并通过 MUSIC 算法识别系统。结束。

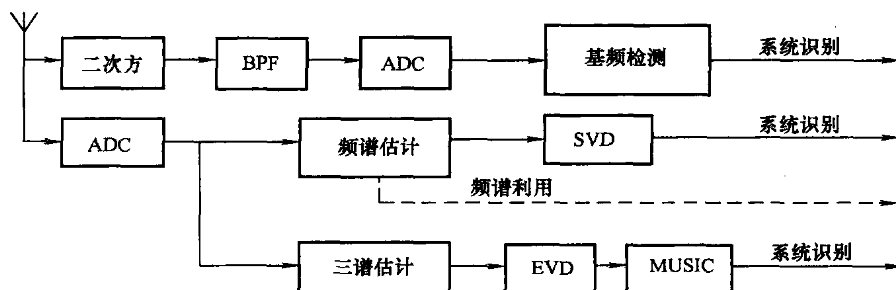


图 7-15 广义多系统感知算法的框图

7.3.1.3 基频

由于数字通信系统中脉冲整形，在符号率（波特率）和其谐波的倒数里有能量尖峰。我们定义具有能量尖峰的最低频率（波特率）为基频。假设 P 个活跃的系统是有基频特点的数字通信系统。传输的第 i 个系统的信号写成如下形式：

$$s_i(t)=\sum_{n=-\infty}^{\infty}x_{i,n}h_i(t-nT_i-\tau_i)\mathrm{e}^{\mathrm{j}(2\pi f_{c,i}t+\alpha_i)} \tag{7-24}$$

对于 $I=1, 2, \cdots, P$，其中 $x_{i,n}$ 是数据序列，$h_i(t)$ 是具有频谱响应 $H_i(\mathrm{jw})$ 的脉冲整形滤波器的脉冲响应，T_i 是符号持续时间，$\tau_i\in[0, T_i)$，$\alpha_i\in[0, 2\pi)$ 分别是时间偏移和相位偏移（在感知过程中都认为是常数），$f_{c,i}$ 是载波频率。具体来说，我们假设 $\{x_{i,n}\}$ 是零均值，方差为 σ_i^2 统计独立并且独立同分布的静态序列。

有很多方法来提取基频信息。这里我们采用非线性频谱线方法。信号先被幅度二次方，然后通过窄带滤波器。假设 $y(t)$ 的实部和虚部有相同的方差，并且和 $w(t)$ 独立。很明显有 $\mathrm{E}\{r^2(t)\}=1/2\cdot\mathrm{E}\{|y(t)|^2+|w(t)|^2\}$。因此，我们得到

$$E\{r^2(t)\}=\frac{1}{2}\sum_{i=1}^{P}\frac{\sigma_i^2\gamma_i^2}{T_i}Z_{i,1}\cos\left[\frac{2\pi(t-\tau_1)}{T_i}\right]+\frac{1}{2}\sum_{i=1}^{P}\frac{\sigma_i^2\gamma_i^2Z_{i,0}}{T_i}+\frac{1}{2}\sigma_w^2 \tag{7-25}$$

其中

$$Z_{i,m}=\frac{1}{2\pi}\int_{-\infty}^{\infty}H_i(\mathrm{j}\theta)H_i^*\left(-\mathrm{j}\left(\frac{2\pi m}{T_i}-\theta\right)\right)\mathrm{d}\theta \tag{7-26}$$

因此信号二次方 $r^2(t)$ 能被分解为

$$r^2(t)=E\{r^2(t)\}+\varepsilon(t)=\frac{1}{2}\sum_{i=1}^{P}\frac{\sigma_i^2\gamma_i^2}{T_i}Z_{i,1}\cos\left[\frac{2\pi(t-\tau_i)}{T_i}\right]+\frac{1}{2}\sum_{i=1}^{P}\frac{\sigma_i^2\gamma_i^2Z_{i,0}}{T_i}+\frac{1}{2}\sigma_w^2+\varepsilon(t) \tag{7-27}$$

式中，$\varepsilon(t)$ 是零均值的扰动项。

我们能观察在式（7-27）中频率为 $\{1/T_i\}$ 的频谱线。在由含有所有可能基频的窄带带通滤波器滤波后，我们能够检测这些频率来识别相应的系统。图 7-16 所示为在 802.11b、802.11g 和蓝牙共存信噪比为 10dB 情况下经过二次方器后频谱的幅度（假设这三个系统有相等的信号能量）。在 1MHz、11MHz 和 20MHz 观察到三个阶跃，这对应于这三个系统的基频。注意，频率幅度主要是由 $Z_{i,1}$ 决定，在相同的信号功率条件下，具有更高基频的系统在通过二次方器之后频率幅度要低一些（假设相同的滚降因子）。另一方面，我们能观察到频带越高，扰动越少。

7.3.1.4 功率谱密度模式

假设在目标频谱内每个备选系统功率谱密度已知，并且被分成 M 个等带宽的

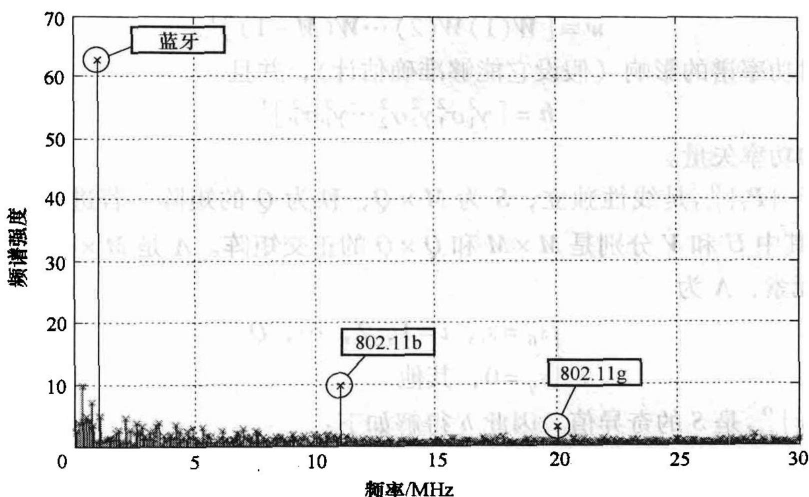


图7-16 通过二次方器后的频谱幅度（$SNR=10\text{dB}$，滚降系数$=0.5$）

子频带。当只考虑功率谱密度形状时，我们将传输的信号考虑成离散时间信号，

$$s_i[n]=u_i[n]*h_i[n]\quad i=1,2,\cdots,P \tag{7-28}$$

式中，$u_i[n]$ 是来自白噪声，它的方差是 σ_i^2。$h_i[n]$ 的频率响应是 $H_i(e^{j\omega})$，这样$\{|H_i(e^{j2\pi k/M})|^2\equiv P_i(e^{j2\pi k/M})\}_{k=0}^{M-1}$是已知实数。因此，接收到的信号的离散时间形式可写成

$$r[n]=\sum_{i=1}^{P}a_is_i[n]+w[n]=\sum_{i=1}^{P}a_iu_i[n]*h_i[n]+w[n] \tag{7-29}$$

令

$$\boldsymbol{p}_i\equiv[P_i(1)P_i(e^{j2\pi/M})\cdots P_i(e^{j2\pi(M-1)/M})]^{\mathrm{T}} \tag{7-30}$$

为第 i 个备选系统的功率谱密度模式。假设这些功率谱密度模式 $\{\boldsymbol{P}_i\}_{i=1}^{Q}$是线性独立的。很多熟悉的谱估计方法，如周期图、Blackman-Tukey 方法、Barlett-Welch 方法、multitaper 法等对不同的频谱估计要求都合适。如果有很长的观察长度可用，那么功率谱估计的误差可以忽略。功率谱估计结果能表示成矢量的形式

$$\hat{\boldsymbol{p}}=\sum_{i=1}^{P}\gamma_i^2\sigma_i^2\boldsymbol{P}_i+\boldsymbol{w}=\boldsymbol{S}\cdot\boldsymbol{h}+\boldsymbol{w} \tag{7-31}$$

其中

$$\hat{\boldsymbol{P}}\equiv[\hat{\boldsymbol{P}}(1)\hat{\boldsymbol{P}}(e^{j2\pi/M})\cdots\hat{\boldsymbol{P}}\ (e^{j2\pi(M-1)/M})]^{\mathrm{T}} \tag{7-32}$$

在频率 $w_k=2\pi k/M$ 的估计是 $\hat{\boldsymbol{P}}(e^{j2\pi(M-1)/M})$，

$$\boldsymbol{S}\equiv[\boldsymbol{p}_1\boldsymbol{p}_2\cdots\boldsymbol{p}_Q]_{M\times Q} \tag{7-33}$$

是功率谱模式矩阵。

$$w \equiv [W(1)W(2)\cdots W(M-1)]^{\mathrm{T}} \tag{7-34}$$

是噪声对功率谱的影响（假设它能够准确估计），并且

$$h=[\gamma_1^2\sigma_1^2\gamma_2^2\sigma_2^2\cdots\gamma_P^2\sigma_P^2]^{\mathrm{T}} \tag{7-35}$$

是接收的功率矢量。

由于 $\{P_i\}_{i=1}^{Q}$是线性独立，S 为 $M\times Q$，秩为 Q 的矩阵。若进行 SVD，$S=U\Lambda V^{\mathrm{T}}$。其中 U 和 V 分别是 $M\times M$ 和 $Q\times Q$ 的正交矩阵。Λ 是 $M\times Q$ 矩阵，其有 $(i,\ j)$ 元素，Λ 为

$$\begin{cases} s_{ii}=s_i,\ i=1,\ 2,\ \cdots,\ Q \\ s_{ij}=0,\ 其他 \end{cases} \tag{7-36}$$

式中，$\{s_i\}_{i=1}^{Q}$是 S 的奇异值。因此 h 得解如下：

$$V\Lambda+U^{\mathrm{T}}\ (\hat{p}-w)=V\Lambda+U^{\mathrm{T}}U\Lambda V^{\mathrm{T}}h=h \tag{7-37}$$

Λ^+是 $Q\times M$ 矩阵，其有 $(i,\ j)$ 元素，Λ^+为

$$\begin{cases} q_{ii}=1/s_i,\ i=1,\ 2,\ \cdots,\ Q \\ q_{ij}=0\quad,\ 其他 \end{cases} \tag{7-38}$$

因此 $\Lambda^+\Lambda=I_Q$。

理想情况下 h 只在对应系统是活跃的情况下包含非零元素。在实际中由于估计误差，h 总是非零。这里运用一种启发式顺序搜索。将 h 的元素按降序排列，$k_1\geqslant k_2\geqslant\cdots\geqslant k_Q$，从 $\hat{p}=1$ 开始计算比值 $\sum_{i=1}^{\hat{p}}K_i/\sum_{i=1}^{Q}K_i$，当这个值超过预先设定的门限时停止。当搜索停止时，则活跃系统的数量由 $\hat{p}$ 决定，相应的元素和系统也将被识别。

图 7-17 所示为按 Welch 的方法得到的功率谱密度的估计。这里假设微波炉的功率谱密度是 2415～2465MHz 的矩形，802.11b 活跃在中心频率 2437MHz。当这些系统的频谱重叠时，估计结果按这些系统的功率谱密度累加。因此，能进行 SVD 来检测系统是否活跃。

7.3.1.5 四阶累积

之前的讨论都基于噪声是协方差矩阵已知的高斯白噪声。不过，如果噪声是协方差未知的有色噪声，要将等式表示成已知的谱矢量的线性组合是不可能的。这就说明测量值的二阶统计量（功率谱密度模式的方法）对于检测和识别多个系统是不够的。这种情况下，高阶统计量成为确保多系统感知成功有用的特征，因为对于高阶统计量，累积量与任何高斯过程无关。若传输的信号 $s_i[n]$ 是式（7-28）的形式，其中 $u_i[n]$ 成为平稳非高斯白随机过程，它的四阶累积量为 ρ_i。$h_i[n]$ 的形式没有变化：p_i 已知并且 $\{P_i\}_{i=1}^{Q}$是线性独立的。$M\times M$ 三谱矩阵 C 的元素定义成

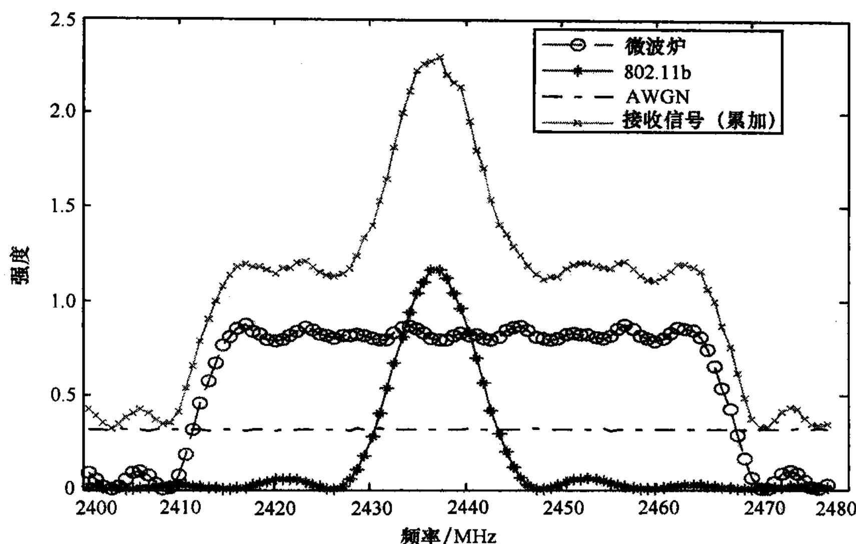


图7-17 微波炉和802.11b（载波频率为2437MHz）的功率谱密度估计

$$c_{ij}=T(\omega_i,\ -\omega_i,\ \omega_j)\quad 1\leqslant i,\ j\leqslant M\quad \omega_i=\frac{2\pi(i-1)}{M} \tag{7-39}$$

这里

$$T(\omega_1,\omega_2,\omega_3)=\sum_{i=1}^{P}\gamma_1^4\rho_i H_i(\omega_1)H_i(\omega_2)H_i(\omega_3)\times H_i(-\omega_1-\omega_2-\omega_3) \tag{7-40}$$

是接收到的三谱矩阵。因此 $\boldsymbol{C}$ 表述为

$$\boldsymbol{C}=\sum_{i=1}^{P}\gamma_1^4\rho_i\boldsymbol{p}_i\boldsymbol{p}_i^{\mathrm{T}}=\boldsymbol{S\Gamma S}^{\mathrm{T}} \tag{7-41}$$

式中，$\boldsymbol{S}$ 是源谱矩阵，$\boldsymbol{\Gamma}$ 是 $Q\times Q$ 的对角矩阵，其非零对角元素的个数为 P。理想情况下，$\boldsymbol{C}$ 是秩为 P 的实矩阵。活跃系统的数量能够通过计算 $\boldsymbol{C}$ 的秩获得。困难的是在实际情况下，由于估计的误差，$\boldsymbol{C}$ 总是满秩的。对三谱矩阵估计进行特征分解，$\hat{\boldsymbol{C}}$ 被写成

$$\hat{\boldsymbol{C}}=\sum_{m=1}^{M}\lambda_m\boldsymbol{g}_m\boldsymbol{g}_m^{\mathrm{H}}=\boldsymbol{G}\sum\boldsymbol{G}^{\mathrm{H}} \tag{7-42}$$

并将特征值按照递减的顺序排列，$\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_M$。

信号子空间通过矢量 $\{\boldsymbol{g}_i\}_{i=1}^{P}$ 展开；噪声子空间通过矢量 $\{\boldsymbol{g}_i\}_{i=p}^{M+1}$ 展开。因此，活跃系统数量的确定和系统本身的识别可以通过顺序搜索或者 MUSIC 算法进行。MUSIC 算法的伪谱计算如下：

$$\overline{R_{\mathrm{music}}}(i)=\frac{\boldsymbol{p}_i^{\mathrm{H}}\boldsymbol{p}_i}{\sum_{m=Q+1}^{M}|\boldsymbol{p}_i^{\mathrm{H}}\boldsymbol{g}_m|^2} \tag{7-43}$$

P 的最大值对应的系统被选中，并识别为活跃系统。

理想情况下，式（7-39）中的三谱矩阵不受叠加的高斯噪声功率的影响。这也意味着四阶累加量的方法可以在低信噪比的环境中表现良好。另一方面，为了对三谱矩阵 C 进行准确的估计，需要很长的收敛时间，这意味着需要更长的感知时间。

7.3.1.6　2.4GHz ISM 频带示例

2.4GHz ISM（工业、科学、医疗）频带是最常见的多个系统共存的环境。这其中还可能包括 WLAN（802.11b 和 802.11g）、蓝牙和微波炉这些潜在的活跃系统。系统的参数列在了表 7-2 中。

表 7-2　系统参数

系　统	载波频率	基　频
802.11b	2412MHz，2437MHz，2462MHz	11MHz
802.11g	2412MHz，2437MHz，2462MHz	20MHz
蓝牙	不固定	1MHz
微波炉	无	无

图 7-18 为这种情况下多系统感知的流程图。如果能量检测显示在频谱中有活跃的系统，我们将尽力锁定在 2412MHz、2437MHz 和 2462MHz 的载波。如果这当中的一些载波被锁定了，那说明 802.11b 或者 802.11g 系统在相应的信道中

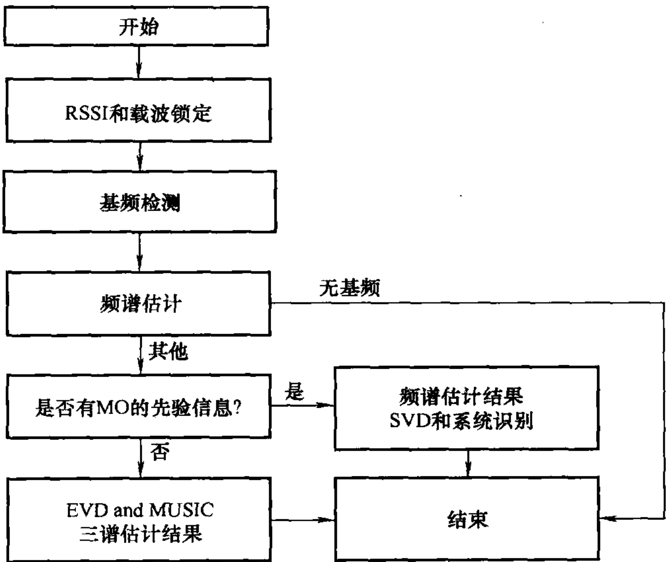


图 7-18　通用多系统感知算法流程图

存在（也可能并存）。若没有锁定任何载波，那说明没有802.11b或者802.11g系统。下一步，采用基频方法。由于以模拟的方式二次方宽带射频频谱需要非常高效的非线性器件，这可能不可实现。在进行感知时，整个频带被分成几个部分。基频方法可以决定802.11b、802.11g和蓝牙系统是否存在，因为它们有不同的基频。

在完成频谱估计后就得到了频谱使用情况，若微波炉的功率谱密度已知，功率谱密度模式方法可以使用。另一方面，若微波炉的功率谱密度未知，将微波炉的干扰看成是加性高斯噪声，四阶累积量方法可以用来识别非高斯信号的其他系统。RSSI基频检测通常比其他方法快，因此不论微波炉的先验信息是否已知，流程图仍然成立，因为感知时间被认为是一个至关重要的参数。

图7-19所示为在$f_1=10\text{MHz}$，$f_2=20\text{MHz}$共存的环境下基频检测的接收机运行特性（ROC）曲线。数据长度为10000，信号经过二次方器和带通滤波器后以100MHz被采样。在图7-20和图7-21是使用功率谱密度模式方法时，检测概率与信噪比的曲线。在表7-2中，考虑三个主要的信道，信道1（载波频率是2412MHz）、信道2（载波频率是2437MHz）、信道3（载波频率是2462MHz）。微波炉也在工作。

7.3.2 无线电资源感知

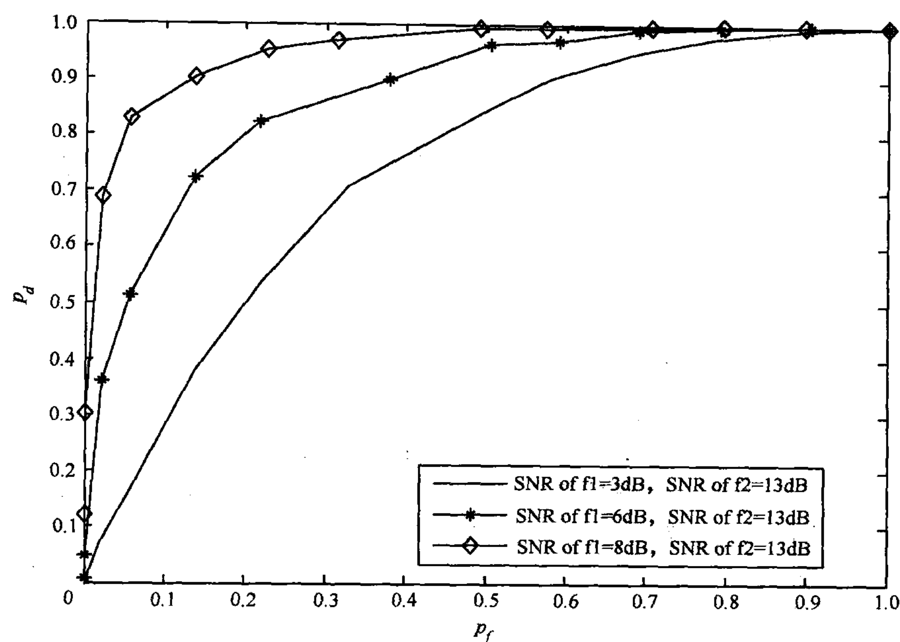
由于通常的频谱感知都是为了在链路层识别认知无线电的传输机会，对于认知无线电网络来说，被动的运行模式下获得的信息是不够的。为了进一步探索多无线电系统的可用性，我们借鉴了一个来自CSMA网络的概念提出了积极感知算法，它与自适用调制编码协同工作。这个算法能决定次要系统是否可以运行，也可以探测在限定的可以承受的干扰下可用的无线电资源。认知无线电在PS可以承受的干扰程度下积极地发射伪载波。通过动态的改变伪载波的功率，并且协作的调整PS的调制和编码参数，认知无线电能决定系统功率的最优使用方案（从而决定可能的速率）。因此，在运行期间的中断和PS的中断概率都能得到控制。次要系统的可用无线电资源能被更有效地得到利用。

7.3.2.1 被动感知的问题

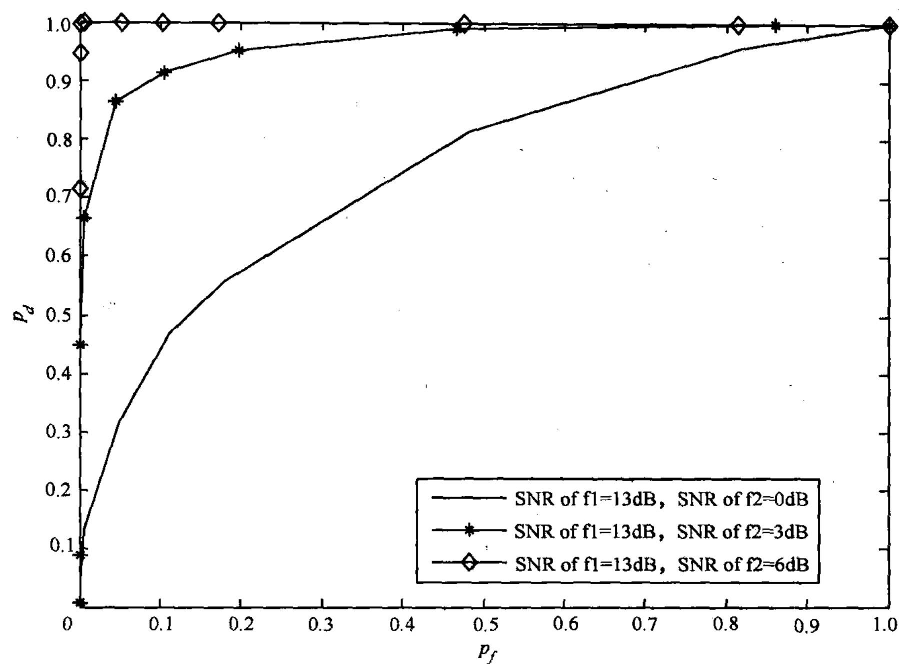
之前介绍的功率缩放控制是一种很自然能想到的方法。但是，它从PS-Tx被动地感知信号，然后推断传输功率控制。这可能存在如下几个问题：

- 最大传输功率的干扰结果可能不准确，这是因为这种方法依赖于关于信道参数α的假设。

- 由于为阴影效应留出的余量是基于整体平均感知的（当PS-Rx和CR-Tx不移动时），PS在感知持续期间的服务质量是不能得到保证的。结果PS-Rx可能存在一段不稳定的也不能被察觉的时间。



a)



b)

图 7-19 基频的接收机运行特性曲线

a) $f_1=10\text{MHz}$ b) $f_2=20\text{MHz}$ 共存

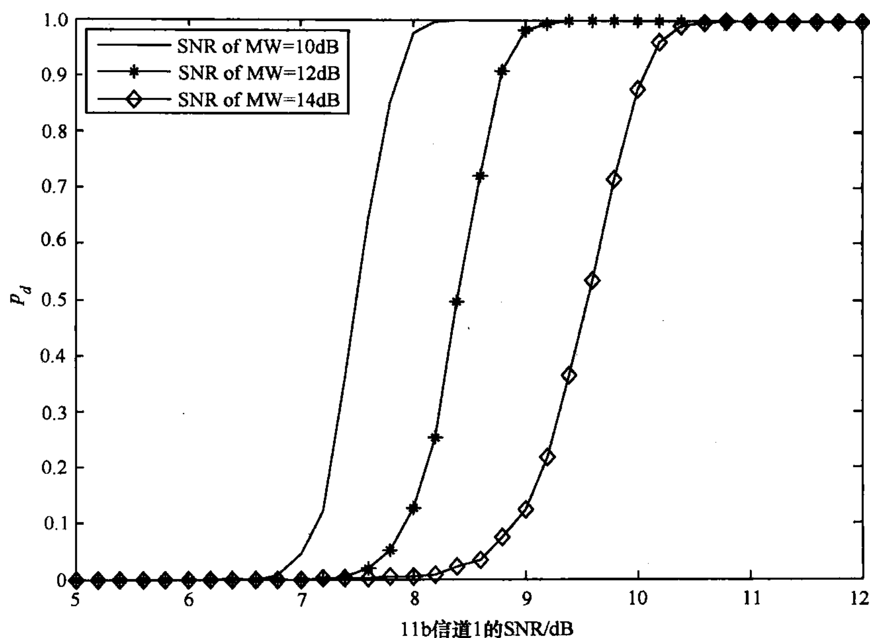


图7-20 检测概率与功率谱密度模式的SNR，11b信道2 = 10dB，信道3 = 10dB

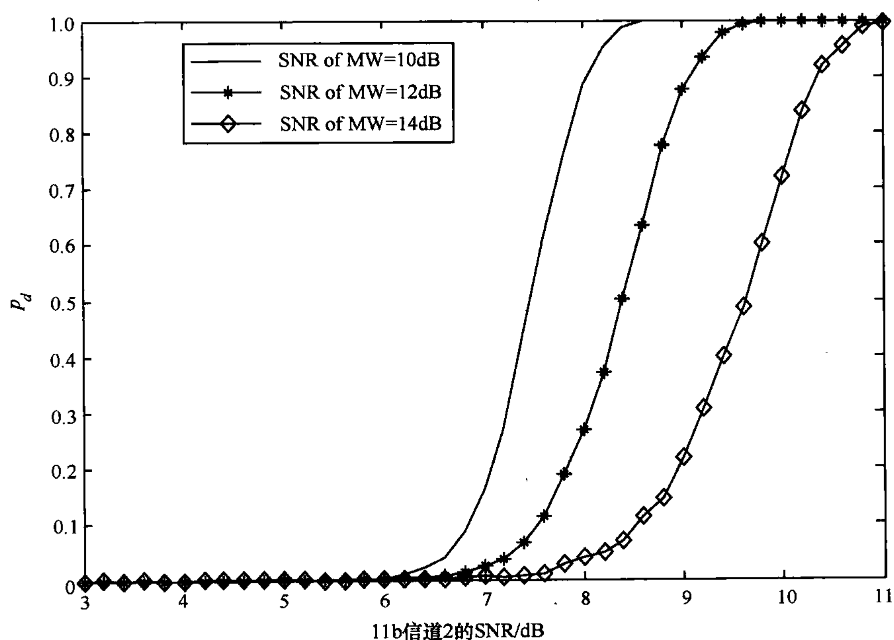


图7-21 检测概率与功率谱密度模式的SNR，11b信道1 = 10dB，信道3 = 10dB

● 次要系统需要保留很大的传输功率余量，用来保护 PS-Rx 的最坏状况。这就可能浪费无线电资源，也就损失了次要系统的系统容量。

尽管协作感知在保护隐藏终端中证明是有用的，但是被动感知的本质没有改变。因此，被动感知的感知行为集中在 PS-Tx 的传输，而不是真正地关注 PS-Rx 的干扰。结果就是最坏情况仍然是被认为是保守地防止 PS-Rx 受到有害干扰，这就导致了相当的无线电资源的减少。依靠被动地监听信道不可避免遇到这些问题，这对有效地使用无线电资源显然不够。主要原因是被动感知完全是在认知无线电端采取的措施。我们介绍一种全新的认知无线电感知的协同机制：积极感知。协同是在认知无线电两端完成的（仅考虑一对认知无线电），已经存在的系统也为受累于隐藏终端问题的次要系统侦测可用的无线电资源。通过协作，认知无线电能真正的测试信道从而知道对已经存在的系统的影响。我们期望积极感知能够解决被动感知存在的问题。此外，积极感知算法能够扩展到认知无线电的多系统环境，这要求传输功率不能干扰任何已经存在的系统。

7.3.2.2 主动探测的问题表述

考虑两对传输的简单的干扰信道模型，一对是具有反馈信道的 PS-Tx 和 PS-Rx，另一对是 CR-Tx 和认知无线电接收机（CR-Rx），图 7-22 为模型图。反馈信道是用来让 PS 交流信道质量和调整传输参数（包括调制方式和编码率），以及优化系统容量。PS-Rx 的 SINR 给出如下：

$$\gamma_{PS} \equiv \frac{P_{PS} \cdot G_{PP}}{P_{CR} \cdot G_{CP} + P_{N,PS}} \tag{7-44}$$

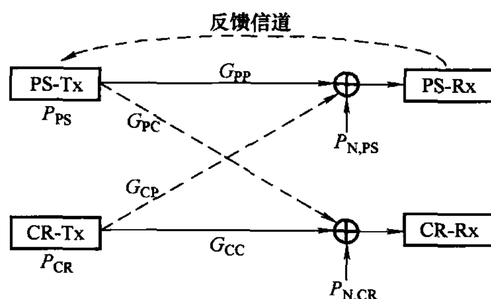


图 7-22 CRN 中拥有主系统反馈信道的干扰信道模型

CR-Rx 的 SINR 为

$$\gamma_{CR} \equiv \frac{P_{CR} \cdot G_{CC}}{P_{PS} \cdot G_{PC} + P_{N,CR}} \tag{7-45}$$

式中，P_{PS}和P_{CR}分别是 PS-Tx 和 CR-Tx 的传输功率，G_{PP}、G_{PC}、G_{CP}和G_{CC}都是信道功率增益，$P_{N,PS}$和$P_{N,CR}$分别是 PS-Rx 和 CR-Rx 的接收功率。

我们考虑慢时变信道，因此信道功率增益 G_{PP}、G_{PC}、G_{CP}和 G_{CC}被认为在感知期间保持不变。对于次要系统，CR-Tx 的传输功率 P_{CR}已知，PS-Tx 的传输功率 P_{PS}可通过主系统和次要系统间的协作获得。此外，不考虑 PS-Tx 的功率控制，因此 P_{PS}是固定的。信道功率的增益 G_{PC}、G_{CC}可通过估计获得，因此被认为已知。值得注意的是作为认知无线电网络里常见的条件，信道功率增益 G_{PC}、G_{CC}对于次要认知无线电系统是未知的。

为了结合 AMC，若 PS-Tx 以调制模式编码 M_j发送一个符号包，例如，M_j对应几种概率：不传输，1/2 率的卷积码 BPSK，1/2 率的卷积码 QPSK，3/4 率的卷积码 QPSK，1/2 率的卷积码 16QAM，3/4 率的卷积码 16QAM，2/3 率卷积码 64QAM 或者 3/4 率卷积码 64QAM。这些被表示为 M_0，M_1，…，M_8。自适应调制编码方法当 $\gamma_{PS} < l_1$ 时，把 PS-Tx 的调制切换到 M_0；$l_1 < \gamma_{PS} < l_2$时，切换到 M_1，等等。注意，当 $\gamma_{PS} < l_1$ 时，PS-Tx 不传输数据，PS 发生中断。在加性高斯白噪声信道下，满足要求的 BER 的唯一调制类型的理论 SINR 如表 7-3 所列。

表 7-3 在 GRAY 映射的 AWGN 信道中，为了满足要求的 BER 而应有的理论 SINR

BER_{req}	BPSK/dB	QPSK/dB	16QAM/dB	64QAM/dB
10^{-2}	4.3	7.3	13.9	19.6
10^{-3}	6.8	9.8	16.5	22.6
10^{-4}	8.4	11.4	18.2	24.3
10^{-5}	9.6	12.6	19.4	25.6

我们考虑两种次要系统的运行限制：

● 在 CR-Tx 的传输期间，PS-Rx 的 SINR 的 γ_{PS}应该大于 l_1，这样才能不对主系统造成中断。

● 为了使接收的信号能够被解码，CR-Rx 的 SINR 的 γ_{CR}大于 η。

这些限制能写成如下形式：

$$\begin{cases} P_{CR} < \dfrac{G_{PP}}{G_{GP}} \cdot \dfrac{1}{l_1} \cdot P_{PS} - \dfrac{P_{N,PS}}{G_{CP}} \\ P_{CR} > \dfrac{G_{PC}}{G_{CC}} \cdot \eta \cdot P_{PS} + \dfrac{\eta}{G_{CC}} \cdot P_{N,CR} \end{cases} \tag{7-46}$$

图 7-23 说明了这些限制。

当

$$\frac{G_{PP}}{G_{CP}} \cdot \frac{1}{l_1} > \frac{G_{PC}}{G_{CC}} \cdot \eta \tag{7-47}$$

时，存在一个可行的区域，在这当中次要系统能够运行。

现在，要确定认知无线电系统是否能够运行，也希望知道对应的次要系统允

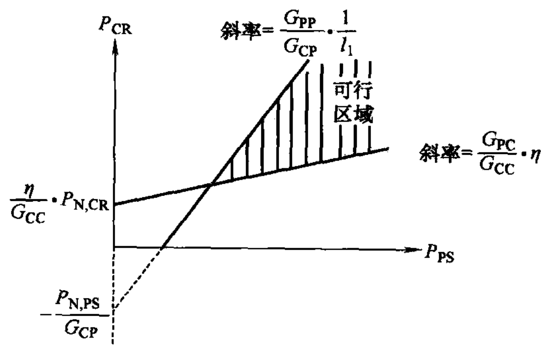


图 7-23 次要系统工作的可行区域

许的最大传输功率。请注意，验证式（7-44）实际上意味着衡量 CR-Tx 给 PS-Rx 的影响。由于 G_{PC}和 G_{CC}已知，l_1 和 η 是预先设定的已知值。可行区域的存在性问题变成了寻找 G_{PP}/G_{CP}的比值。

7.3.2.3 具有协作 AMC 的积极感知算法

借鉴 CSMA 在传输前“测试”信道是否为空的概念，CR-Tx 能够通过测试感知信道来评估 PS-Rx 的干扰。这种频谱感知概念被称为积极感知（或者主动感知），这与通常的被动感知相对。在积极感知中，CR-Tx 通过以下方式感知信道：(i) 发射动态控制功率的伪载波；(ii) 观察拥有反馈信道的 PS 传输行为的变化。因为传输行为的变化是由于动态控制的伪载波的干扰造成的。最终推导出 CR-Tx 的传输对 PS-Rx 的影响。

协作 AMC 包括三个步骤：

1. PS-Tx 根据 PS-Rx 的 SINR 调整调制和编码参数（MCP）来维持要求的比特错误率。

2. PS-Tx 与次要系统合作，使传输功率和 MCP 等 PS-Tx 传输参数能提供给次要系统。

3. 次要系统调整自己的传输功率 P_{PS}，适应调制和编码策略来控制 PS-Rx 的干扰，同时最优化使用无线电资源。由于 PS-Tx 和次要系统的协作，主系统的 MCP 和 P_{PS}次要系统都可以获取。这个协作可能不需要修改主系统，因此能够通过解码控制信息（例如帧头）在认知无线电上实现，或者通过收到信号的检测过程。

结合积极感知和协作 AMC 的概念，我们提出一种处理隐藏终端的算法。这个算法总结如下，初始值为 $k=1$，$I_{CR}[1]=\Delta I_{CR}[0]$：

1. 检测 MCP。若 MCP 与 SINR 水平相当，仅高于无传输，则结束；否则，转到步骤 2。

2. 传输功率为 $I_{CR}[k]$ 的伪载波。

3. 等待时间 τ。

4. 检测 MCP，若对应的 SINR 和 MCP 相比变得比步骤 1 中的低一档，转到步骤 6，若减少不止一个档位，结束。否则，增加伪载波的功率 $\Delta I_{CR}[k]$。

5. 如果 $I_{CR}[k] \geqslant P_{regulation}$，结束。否则，把 k 增加 1，然后转到步骤 3。

6. 用 γ_{switch} 代表对应的 SINR 转换档，如果

$$\frac{\gamma_{switch}}{l_1} \cdot \frac{\Delta I_{CR}[k]}{P_{PS}} > \frac{G_{PC}}{G_{CC}} \cdot \eta \tag{7-48}$$

成立，则次要系统的运行可行；否则，不行，结束。

图 7-24 给出了相应的流程图。在第 1 步，对应于最低传输率的 SINR 是算法结束，这是为了避免由于伪载波的传输导致 PS 中断。在第 5 步，当伪载波的功率超过给次要系统规定的传输功率 $P_{regulation}$，这意味着次要系统在最大允许传输功率 $P_{regulation}$ 下可以通信，并且不对 PS-Rx 造成不能承受的干扰。在第 6 步，转换 SINR γ_{switch} 表达成

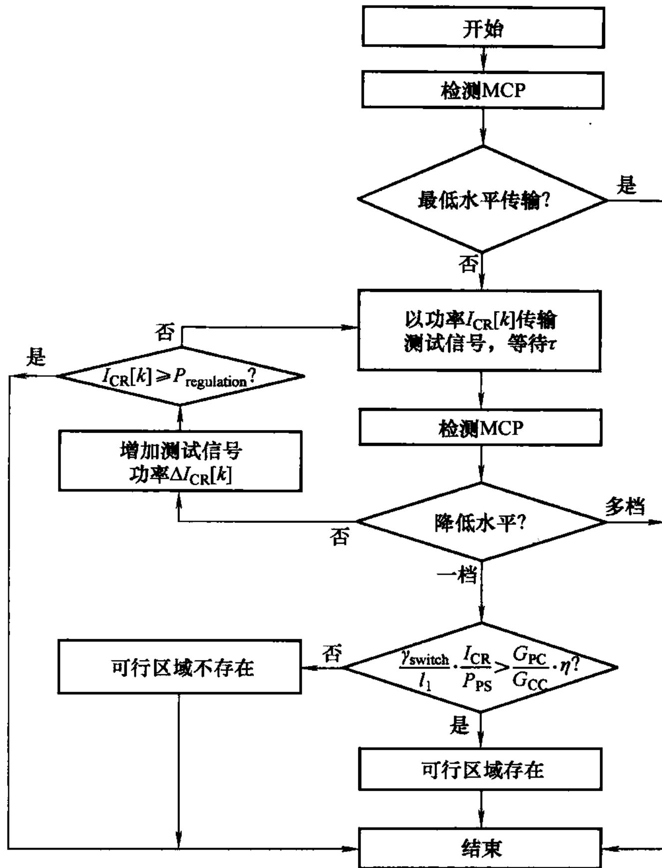


图 7-24　积极感知算法的流程图

$$\gamma_{switch} = \frac{P_{PS} \cdot G_{PP}}{I_{CR} \cdot G_{CP} + P_{M,PS}} \tag{7-49}$$

基于干扰功率占绝对优势的假设，γ_{switch}可以近似为

$$\gamma_{switch} \approx \frac{P_{PS} \cdot G_{PP}}{I_{CR} \cdot G_{CP}} \tag{7-50}$$

从而功率增益比计算得

$$\frac{G_{PP}}{G_{CP}} \approx \gamma_{switch} \cdot \frac{I_{CR}}{P_{PS}} \tag{7-51}$$

通过将式（7-51）代入式（7-47），可以推导出式（7-48）。因此，当式（7-48）成立，次要系统的运行是可行的。根据式（7-51），CR-Tx 允许的最大的传输功率 $P_{max,CR}$可表述为

$$P_{max,CR} = \frac{G_{PP}}{G_{CP}} \cdot \frac{P_{PS}}{l_1} \tag{7-52}$$

注意，在这种情况下，P_{PS}通过协作 AMC 被认为对于认知无线电是已知的。若 P_{PS}未知，积极感知算法通过两次检测 PS-Tx 的 MCP 变量，使其仍然能够工作。

7.3.2.4 调制和编码参数检测

通常，MCP 的检测能够通过以下两种方法中的一种完成：（i）MCP 信息能直接从译码后的导频或者帧头中获得；（ii）MCP 信息能够通过感知接收到的数据信号检测获得。在第一种方法中，可以使用 PS 传统的信号解码步骤。但是，在这里只考虑第二种方法以保证积极感知算法的可靠性和通用性。MCP 被分解成两部分：调制类型和编码率。调制类型的检测作为调制分类问题已经被很好的研究，它是信号检测/截获和信号解调的中间步骤。另一方面，在解码后，从数据信号中检测解码率还没有大家熟知的好方法。我们介绍一种检测方法，这种方法针对有同步字的帧的结构特征。

通常来说，有两种调制分类算法，包括决策-理论/最大似然比方法和基于特征的方法。接收到的信号表示为

$$r(t) = s(\mu_i; t) + n(t) \quad 0 \leqslant t \leqslant KT_s \tag{7-53}$$

式中，T_s 是符号持续期，K 是观察到符号数，$n(t)$ 是双边功率谱密度为 $N_0/2$ 高斯白噪声、包括接收机噪声、干扰和人为干扰。具体地，

$$s(\mu_i; t) = e^{j\theta} e^{j2\pi\Delta f t} \sum_{k=1}^{K} s_k^{(i)} e^{j\varphi_k} g(t - (k-1)T_s - \varepsilon T_s) \tag{7-54}$$

是收到信号无噪声的基带复包络。θ 是载波相位，Δf 为频率偏移，$S_k^{(i)}$ 是在区间 $[(k-1)T_s,\ kT_s]$ 的传输的数据符号，φ_k 是相位抖动，ε 是时间偏移，$g(t)$ 是发射机脉冲形状和信道冲击响应的卷积。μ_i 表示接收到信号与未知符号和信号

参数之间的依赖度。接着，我们将考虑部分这些未知量，而认为其他未知量对于接收机是已知，这样就简化了接收机的复杂度。

在决策-理论方法中，调制分类被表述成多重复合假设检验问题。通过比较似然比和预定的门限来做判断。随着不同的关于未知量的假设，开发了似然比检验的变体。平均似然比检验（ALRT）把未知量当做随机变量，并假设它们的概率密度函数已知。只要我们假定的概率密度函数和真实的一致，ALRT 能保证最优解。但是，大多数情况下，ALRT 通常比较复杂，因此，它被简化为准对数似然比（qLLR）。这种通用化的似然比（GLRT）检测把未知量当成确定量，这些确定量在认为每个假设为真时可以估计得到。GLRT 和 qLLR 相比，在区分 16QAM 和 V. 29 时提供更小的错误概率。但是，GLRT 不能区分 BPSK 和 QPSK。复合似然比检测（HLRT）是用 ALRT 来处理一些未知量，用 GLRT 处理另外的未知量。我们运用 HLRT 将数据序列当成具有均匀概率随机变量处理，载波相位仍然是未知的确定量来区分 BPSK 和 QPSK。

总的来说，基于特征的系统包括两个阶段：特征提取和模式识别。特征在提取后被用来将收到的信号划分到不同的调制类型。直觉上，信号本身和它的瞬时幅度、频率和相位在时间域提供了一组很好的特征，当然即使过零也可以使用。最具有描述性的特征毫无疑问是 PDF，但是这不容易获得，因此矩和累积量，矩的变化成为第二选择。如果频谱是非平稳的，可以使用小波变换。在次要系统的运行环境里，调制分类通常是在预先确定的主系统库中决定调制的类型。因此，预先选择近似分类法。

空对地传输中数据-链路协议最常采用已知二元序列或者同步字来实现合适的帧同步。这类特点可能被用作决定信道编码的细节。例如，若我们对接收的信号采取 1/2 率卷积码的 Viterbi 解码，并且如果解码后的序列出现了这些同步字中的一个，很可能传输信号是以这种方式编码的。对于 MCP 检测，调制类型和编码率通过使用各种可能组合联合决定。假设收到信号是被可能调制类型中的某一种调制，然后用可能的编码率中的一种进行编码。通过在这样的假设下解调和译码，我们能搜索每一个序列来寻找同步字序列。若发生这种情况，调制类型和编码率就确定了。

7.3.2.5　伪载波设计

伪载波的目的是动态改变 PS-Rx 受到的干扰，这同时也改变了主系统的 MCP 受到的干扰。当设计伪载波的机制时，MCP 检测周期 T_D 和增加功率的程度 $\Delta I_{\mathrm{CR}}[k]$ 是两个至关重要的参数。定义 P_{target} 为伪载波用来改变 MCP 的最终目标功率水平。$K=\left\lfloor\frac{P_{\mathrm{target}}}{\Delta I_{\mathrm{CR}}[k]}\right\rfloor$是 $\Delta I_{\mathrm{CR}}[k]$ 比 P_{target} 低的最大周期。其中 $\lfloor\cdot\rfloor$ 是向下取整函数。T_D 和 $\Delta I_{\mathrm{CR}}[k]$ 的关系表示为

$$I_{CR}(t)=I_{CR}[k],(k-1)T_D\leqslant t\leqslant kT_D \tag{7-55}$$

并且

$$I_{CR}[k+1]=I_{CR}[k]+\Delta I_{CR}[k] \tag{7-56}$$

对于 $k=K+1$，伪载波功率超过 P_{target}，而且从伪载波传输到 MCP 改变的瞬时延时为 $\tau_{transient}$，

$$\tau_{transient}=\tau_{SINR,PS}+\tau_{feedback,PS}+\tau_{adaption,PS}+\tau_{detection,CR}=\tau+\tau_{detection,CR} \tag{7-57}$$

式中，$\tau_{SINR,PS}$是 PS-Rx 的 SINR 估计要求的时间；$\tau_{feedback,PS}$是反馈信息传递的延时；$\tau_{adaption,PS}$是 PS-Tx 根据反馈信息改变的 MCP；$\tau_{adaption,CR}$是次要系统要求的 MCP 检测时间。T_D 应该比总的延时 $\tau_{transient}$ 长，以此来保证算法的成功。但是 T_D 不能太长，因为之后保持在不足的伪载波功率水平的周期变长。之后我们仍将讨论这点。

增加的功率量 $\Delta I_{CR}[k]$ 可以是预先设定的或者自适应的：预先设定的方法通过一个确定函数来设定 $\Delta I_{CR}[k]$，这个确定函数可以通过实验来获得。自适应的方法根据当前的环境来设定 $\Delta I_{CR}[k]$。为了简化分析，这里只考虑预先设定的方法，并且以函数为常数作为例子，即 $\Delta I_{CR}[k]=\Delta I_{CR}$是常数。如图 7-25 所示，这受限于两个因素之间的折中：误差度和能量损失。当 $\Delta I_{CR}[k]$ 太大时，误差度可能很严重。因此，最终的功率水平 $\Delta I_{CR}[K+1]$ 可能会有巨大的误差，或者导致在算法的第 4 步中 MCP 的多级下降，这将导致 G_{PP}/G_{CP}的估计误差。另一方面，能量损失可以计算如下：

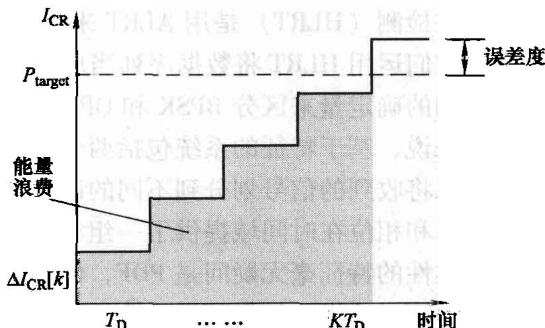


图 7-25 误差度和能量浪费示例

$$E_{waste}=\sum_{k=1}^{K}\Delta I_{CR}\cdot kT_D=\frac{K(K-1)}{2}\Delta I_{CR}\cdot T_D\propto\frac{T_D}{\Delta I_{CR}} \tag{7-58}$$

为了减少 E_{waste}，T_D 应该小，而对于一定 P_{target}，$\Delta I_{CR}[k]$ 应该大。注意 T_D 的下限为 $\tau_{transient}$，由于误差度的缘故，$\Delta I_{CR}[k]$ 不应该太大。因此在设计伪载波时存在折中。

7.3.2.6 扩展到多系统环境

让我们回到最初的应用场景：可能有多个主系统，在所有节点的接收机，次要系统不允许造成毁坏性的干扰。假设有 M 个主系统，每个主系统有一个传输对。PS-Rx 在第 m 个主系统的 SINR 表示成

$$\gamma_{PS,m}\equiv\frac{P_{PS,m}\cdot G_{Pm,Pm}}{P_{CR}\cdot G_{CR,Pm}+\sum_{i=1,i\neq m}^{M}P_{PS,i}\cdot G_{Pi,Pi}+P_{N,PSm}} \tag{7-59}$$

在多主系统的情况下，最大允许传输功率 $P_{max,CR}$ 必须保证每个 PS-Rx 的 SINR 都高于 l_1，即

$$\gamma_{PS \cdot m} > l_1, m = 1, \cdots, M \tag{7-60}$$

我们可以积极地感知信道并锁定一个特定主系统（假设是第 i 个主系统）来观察它的传输参数的变化，然后计算在这个主系统限制下最大允许的传输功率，记作 $P^i_{max,CR}$。这个步骤直到对于所有的主系统 $P^i_{max,CR}$ 都确定才停止。最大允许的传输功率是

$$P_{max,CR} = \min\{P^1_{max,CR}, \cdots, P^M_{max,CR}\} \tag{7-61}$$

但是，积极感知过程和一个主系统的情况稍有不同：传输参数的变化应该被观察两次。针对第 m 个主系统的第一次转换 SINR 为

$$\begin{aligned}\gamma^1_{switch} &= \frac{P_{PS,m} \cdot G_{Pm,Pm}}{P^1_{CR} \cdot G_{CR,Pm} + \sum_{i=1,i\neq m}^{M} P_{PS,i} \cdot G_{Pi,Pi} + P_{N,PSm}} \\ &= \frac{P_{signal}}{P^1_{CR} \cdot G_{CR,Pm} + P_{I,PS} + P_{N,PSm}}\end{aligned} \tag{7-62}$$

式中，$P_{PS,m} \cdot G_{Pm,Pm} = P_{signal}$，$\sum_{i=1,i\neq m}^{M} P_{PS,i} \cdot G_{Pi,Pi} = P_{I,PS}$

第二次转换 SINR 为

$$\gamma^2_{switch} = \frac{P_{signal}}{P^2_{CR} \cdot G_{CR,Pm} + P_{I,PS} + P_{M,PSm}} \tag{7-63}$$

CR-Tx 的达到主系统限制的传输功率为

$$\frac{P_{signal}}{P^m_{max,CR} \cdot G_{CR,Pm} + P_{I,PS} + P_{N,PSm}} = l_1 \tag{7-64}$$

已知的参数包括 γ^1_{switch}、P^1_{CR}、γ^1_{switch}、P^2_{CR}、l_1，还需解出 P^t_{CR}。在一些推导之后，得到

$$P^m_{max,CR} = \frac{\gamma^2_{switch} P^2_{CR}(\gamma^1_{switch} - l_1) + \gamma^1_{switch} P^1_{CR}(l_1 - \gamma^2_{switch})}{l_1 \cdot (\gamma^1_{switch} - \gamma^2_{switch})} \tag{7-65}$$

7.3.2.7 实例

为了说明积极感知（或主动感知），考虑次要系统的理想条件拥有如下假设（见图 7-26）：

- PS-Tx 与 PS-Rx 间的反馈信道为理想信道。
- P_{PS} 与 MCP 可以通过主系统与次要系统间的协作准确获得。
- G_{PC} 与 G_{CC} 可通过次要系统准确估计得到。
- P_{target} 可无误差达到。
- 噪声功率小于干扰功率，从而考虑 SINR 时可忽略噪声功率。

基于以上假设，在确定 $P_{max,CR}$ 时，主要误差来源于 PS-Rx 的 SINR 估计。不

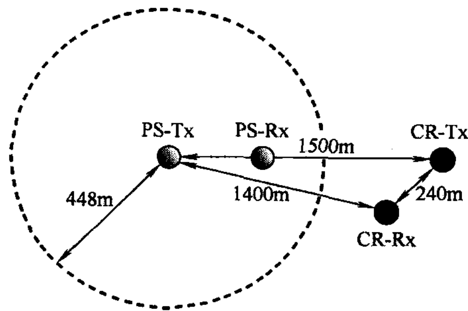


图 7-26 仿真环境

失一般性，PS-Rx 的 SINR（dB）估计值可建模为服从高斯分布的随机变量 $\hat{\gamma}$，其中 $E\{\hat{\gamma}\}=\gamma$ 为实际的 SINR 值，σ_{SINR} 为标准差。$P_{max,CR}$ 重写为

$$P_{max,CR}(dBm)=\hat{\gamma}_{switch}(dB)-\gamma_{switch}(dB)+[G_{PP}(dB)-G_{PC}(dB)]+P_{PS}(dBm)-l_1(dB) \tag{7-66}$$

式中，$\hat{\gamma}_{switch}$ 为 PS-Rx 的转换 SINR 估计值。

定义 PS 的中断概率为

$$\begin{aligned}P_{outage,PS}&=Pr\{\gamma_{PS}<l_1(dB)\}\\&=Pr\{P_{PS}(dBm)-P_{CR}(dBm)+[G_{PP}(dB)-G_{PC}(dB)]<l_1\}\end{aligned} \tag{7-67}$$

在算法中，$P_{CR}=\delta P_{max,CR}$（线性尺度），其中 δ 用来为关于 $P_{outage,PS}$ 的限制保留余量，于是有

$$\begin{aligned}P_{outage,PS}&=Pr\{\gamma_{switch}(dB)-\hat{\gamma}_{switch}(dB)<10\log\delta\}\\&=Q\left(\frac{10\log\delta}{\sigma_{SINR}}\right)\end{aligned} \tag{7-68}$$

请注意，$P_{outage,PS}$ 受 δ 和 PS-Rx 的 SINR 估计的准确度影响，但是不受传播模型的参数影响，因为我们的算法实质上是感知在 PS-Rx 的干扰。

作为比较，请回顾通常功率缩放控制技术的最优情况。假设 PS-Tx 和 CR-Tx 之间的距离能够被准确估计，那么认知无线电就能确定最大传输功率。把传播模型考虑成一个路径损失模型，因此信道功率增益写成

$$G(r)=r^{-\alpha} \tag{7-69}$$

式中，r 是发射机和接收机之间的距离，α 是路径损失的指数。假设系统参数是 $P_{PS}=10dBm$，$\Delta=71.6dB$，$u=1dB$，$\sigma^2=-70dBm$，$d=1500m$，$\gamma_{dec}=l_1=8.4dB$。图 7-26 所示为 PS-Rx 的 SINR 和 CR-Tx 的传输功率的关系。在 PS-Rx 的 SINR 限制下，CR-Tx 的最大传输功率随不同环境而不同。因此，当对环境的假设和实际情况不符时，会产生对主系统的有害干扰。表 7-4 总结了不同环境的 α 值。图 7-27 中的结果能转换到图 7-28 中关于主系统的中断概率。若一个

PS-Rx在主系统的服务范围内（SINR $\geq \gamma_{dec}$），当次要系统信号没有出现且SINR$\geq\gamma_{dec}$。当次要系统信号出现，PS-Rx 发生中断。中断概率定义为由次要系统造成的中断百分比（假设 PS-Rx 是均匀分布的）。图 7-29 给出了对于通常频谱感知主系统的等效中断概率。

表 7-4　典型路径损失指数

环　　境	α 范围
城市大蜂窝	3.7～6.5
城市小蜂窝	2.7～3.5
办公楼（同一层）	1.6～3.5
办公楼（多层）	2～6
商店	1.8～2.2
工厂	1.6～3.3
家庭	3

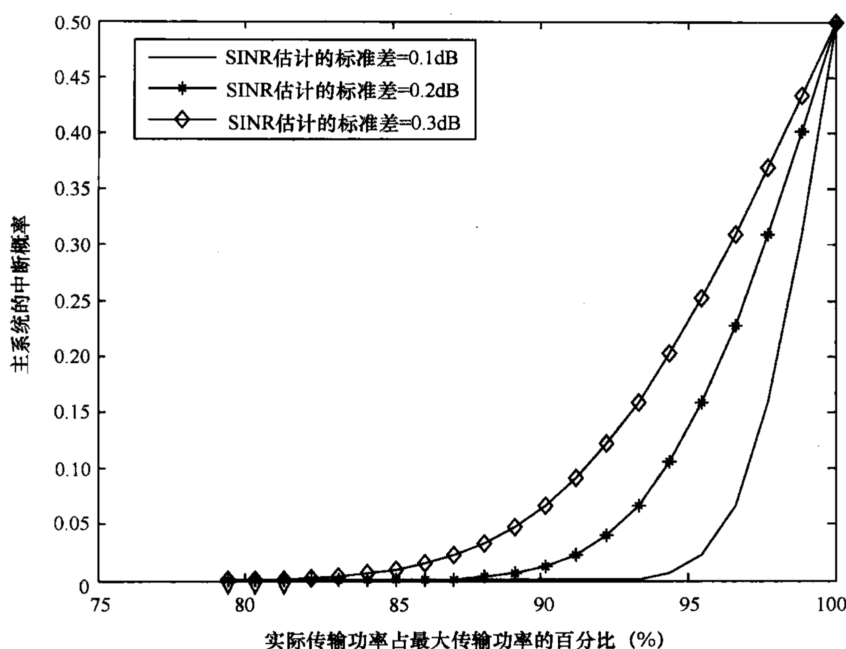


图 7-27　积极感知算法中主系统的中断概率

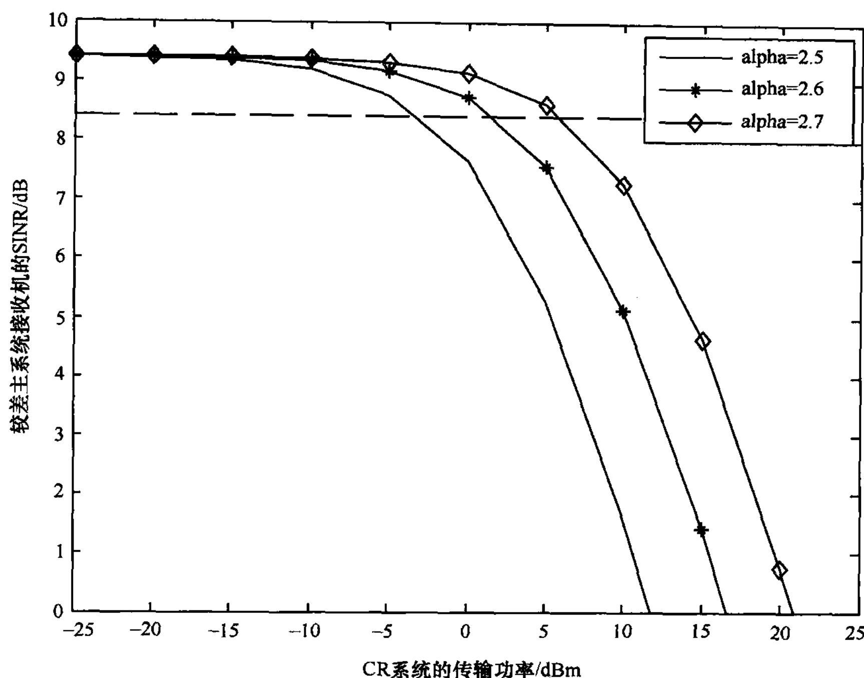


图 7-28　通常感知中主系统接收机最坏情况的 SINR

在图 7-28 中考虑 $\alpha=2.7$。在 $\gamma_{dec}=8.4$ 的限制下，通常感知中 $P_{CR}=6210.7\text{dBm}$。在图 7-29 中，相应的中断概率为 $P_{outage,PS}=0.02$。假定 $\sigma_{SINR}=0.3\text{dB}$，$P_{outage,PS}=0.02$，在积极感知算法中相应的 $\delta=0.8675$。考虑图 7-30 中所示的具体仿真环境，在 PS-Rx 一侧，图 7-30 所示为当 PS-Rx 沿着从 PS-Tx 到 CR-Tx 线运动时的平均 SINR。位置 0m 对应 PS-Tx。根据这些参数，在图 7-30 中可以观察到当位置大于 410m 时，在通常感知中 PS 的中断发生，而在我们的算法中没有发生中断。只要 PS-Rx 不向 PS-Tx 运动，通常感知中的中断一直持续。这显示了长时间通信的中断问题和我们算法的改进。如图 7-31 所示，次要系统归一化的容量可以推导为

$$\frac{C}{B}=\log_2\left(1+SINR\right) \tag{7-70}$$

运用积极感知算法，完成了可观的容量改善。因为积极感知算法实际上是感知到 PS-Tx 的信道，而不是考虑更坏情况。注意，在位置大于 410m 时，通常感知的容量比积极感知容量要高，因为不能承受的 P_{CR} 造成了 PS 的中断。

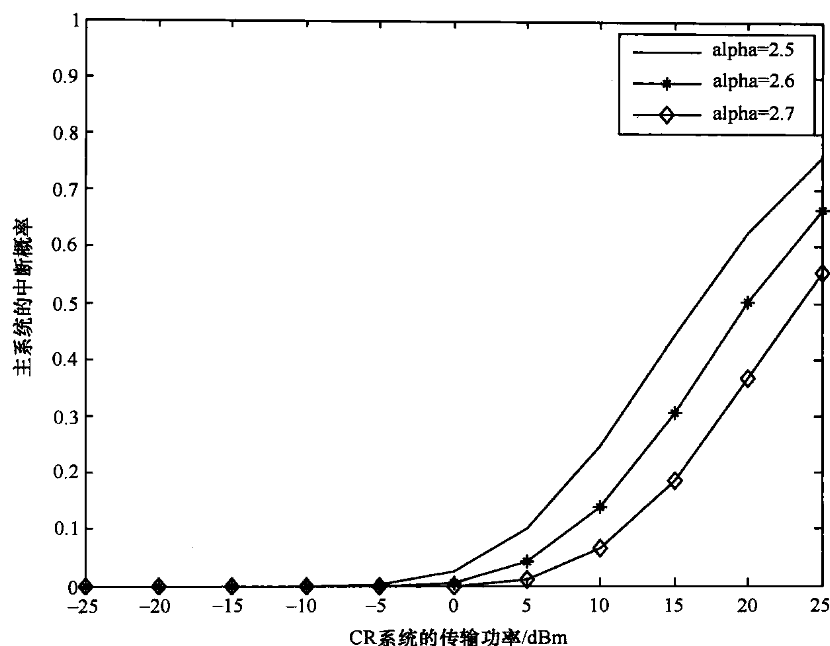


图 7-29　通常感知中主系统的等效中断概率

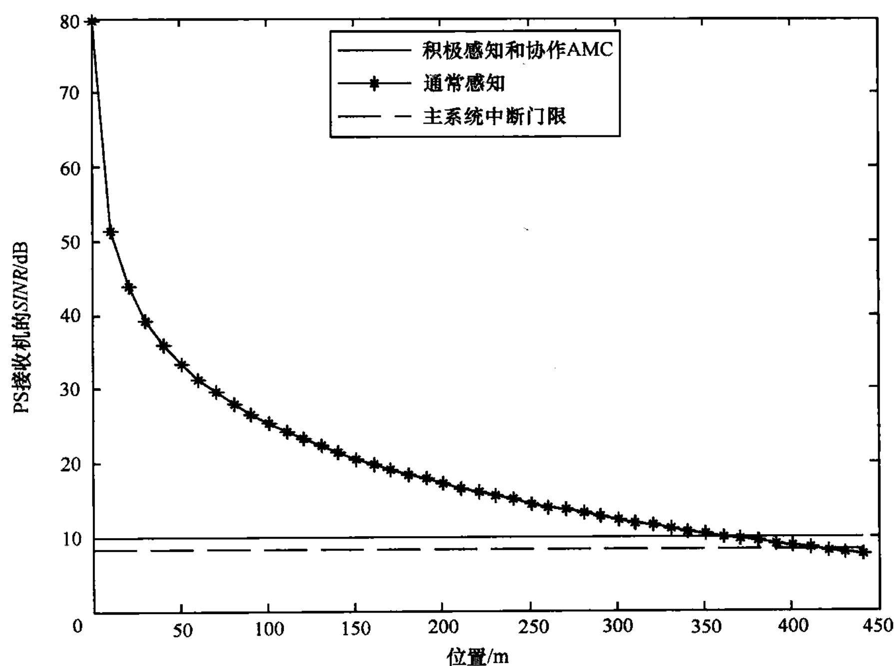


图 7-30　PS-Rx 位置及其对应的 *SINR*，$\delta=0.8675$，$P_{CR}=6.1207\text{dBm}$

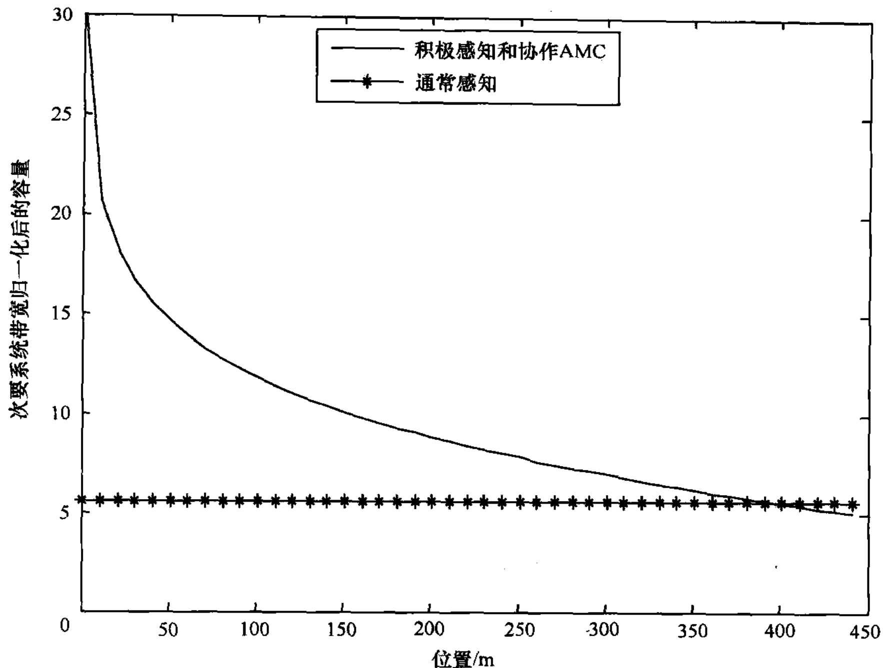


图 7-31 PS-Rx 位置及其对应的次要系统的归一化容量，$\delta=0.8675$，$P_{CR}=6.1207\text{dBm}$

参考文献

[1] S.M. Mishra, A. Sahai, R.W. Brodersen, 'Cooperative Sensing Among Cognitive Radios', in *Proc. IEEE Int. Conf. Communications*, **4**, pp. 1658–1663.

[2] E. Visotsky, S. Kuffner, R. Peterson, 'On Collaborative Detection of TV Transmissions in Support of Dynamic Spectrum Sharing', in *Proc. 1st IEEE Int. Symp. New Frontiers in Dynamic Spectrum Access Networks DySPAN*, 2005, pp. 338–345.

[3] M. Gudmundson, 'Correlation Model for Shadow Fading in Mobile Radio Systems', *Electron. Lett.*, **27**(23), 1991, 2145–2146.

[4] C-X Wang, H-H Chen, X. Hong, M. Guizani, 'Cognitive Radio Network Management', *IEEE Vehicular Technology Magazine*, **3**(1), 2008, 28–35.

[5] J. Unnikrishnan, V.V. Veeravalli, 'Cooperative Sensing for Primary Detection in Cognitive Radio', *IEEE Journal on Selected Topics in Signal Processing*, **2**(1), 2008, 18–27.

[6] Z. Quan, S. Cui, A.H. Sayed, 'Optimal Linear Cooperation for Spectrum Sensing in Cognitive Radio Networks', *IEEE Journal on Selected Topics in Signal Processing*, **2**(1), 2008, 28–40.

[7] A. Ghasemi, E.S. Sousa, 'Interference Aggregation in Spectrum Sensing Cognitive Wireless Networks', *IEEE Journal on Selected Topics in Signal Processing*, **2**(1), 2008, 41–56.

[8] G. Ganesan, J. Li, 'Cooperative Spectrum Sensing in Cognitive Radio, Part I: Two-User Networks', *IEEE Transactions on Wireless Communications*, **6**(6), 2007, 2204–2213.

[9] G. Ganesan, J. Li, 'Cooperative Spectrum Sensing in Cognitive Radio, Part II: Multiuser Networks', *IEEE Transactions on Wireless Communications*, **6**(6), 2007, 2214–2222.

[10] S.Y. Tu, K.C. Chen, R. Prasad, 'Spectrum Sensing of OFDMA Systems for Cognitive Radio Networks', to appear in the *IEEE Transactions on Vehicular Technology*, 2009.

[11] C.K. Yu, K.C. Chen, 'Multiple Systems Sensing for Cognitive Radio Networks over Rayleigh Fading Channel', *Proceedings IEEE VTC Spring, Singapore*, 2008.

[12] C.K. Yu, Multiple System Sensing for Cognitive Radio Networks, Master Thesis, National Taiwan University, 2008.

第 8 章　媒体接入控制

认知无线电最简单的概念（认知无线电网络的最低两层）是利用频谱感知去检测传输机会，并动态地接入频谱（DSA）。讨论过认知无线电网络中的频谱感知后，我们将介绍认知无线电网络中的媒体接入控制的更多信息，关于协调多认知无线电以伺机的方法接入共享的媒体，这不同于在链路层为伺机的传输而采用的简单频谱使用。

8.1　认知无线电的媒体接入控制

首先我们介绍媒体接入控制操作的一般数学模型。注意 lai、El Gamal、Jiang 和 poor 在参考文献［35］中从统计通信理论推导出最优的媒体接入方法。

虽然可以考虑多维频谱资源，我们还是先考虑一个频带的集合作为一般情况。假设感兴趣的频带（典型地，主系统操作）是一个数字标记的集合，$M=\{1, \cdots, M\}$。在时间 t_n，认知无线电网络操作进行一次频谱使用更新。第 n 次观察（或者位置）的时间间隔是 $\{t_n, t_{n+1}\}$。由于每条链接的伺机的本质（即频段）被建模成一个马尔科夫链，第 i 个频带可用的概率是一个概率为 π_i的、时不变的伯努利过程。

定义如下指示函数（类似于第 5 章中定义的清晰的信道的指示函数）：

$$\mathbf{1}_i[n]=\begin{cases}1,[t_n,t_{n+1}]\text{上可用的频带 } i\\0,\text{其他}\end{cases}\tag{8-1}$$

对于理想的频谱感知，我们能以可靠的方式决定式（8-1）。然而，任何频谱感知都有一些易受攻击的情况，因此我们需要考虑更多因素决定媒体接入控制。接着式（8-1），伯努利随机变量的概率群分布函数（pmf）在第 i 个频带表示为

$$f_{\mathbf{1}_i[n]}(x\mid\pi_i)=\pi_i x+(1-\pi_i)\delta(x)\tag{8-2}$$

此处，可以合理的假设 $\{\mathbf{1}_i\ [n]_{i=1}^{M}\}$ 是独立的，其中 $n=1, \cdots, L$，L 表示观察间隔深度。记为，

$$\pi=[\pi_1,\cdots,\pi_M]$$

对于可靠的认知无线电操作，频谱感知是必要的，因此 CR-Tx 可以获得每个频带的可用信息。但是，在认知无线电链接之上网络操作，策略将高度依赖于 π。

情况1：π 已知。

情况2：π 未知。

情况3：π 可以通过一些认知无线电网络感知或者断层扫描方法（网络断层扫描将在第9章介绍）探测或者估计。

传统认知无线电函数如下所示：在时间 t_n，认知无线电获悉一个选定频带 s_n 可用（通过频谱感知）。如果 $1_{s_n}(n)=1$，在带宽 B 内的信息可被成功传输。对于 L 时间段内，总吞吐量为

$$W=\sum_{n=1}^{L}1_{s_n}(n) \tag{8-3}$$

当 π_i 已知的情况下，对于认知无线电的频谱感知策略是简单地挑选信道 $\hat{i}=\mathrm{argmax}_{i\in M}\pi_i$ 去感知。基于特定的决策标准和条件，如部分观察的马尔科夫决策过程[27]，接入决策可被最优化或次最优化地制定。本章的意图不在于探求数学和理论解，而将探索一些基本的认知无线电中媒体接入（或者 DSA），即复合认知无线电功能。

8.2 多信道媒体接入控制

在共存的多无线电系统的环境中，认知无线电中的媒体接入控制也许为获得多信道而竞争，这使得多信道媒体接入控制（McMAC）的研究引起注意，即使几乎没有关于 McMAC 在认知无线电网络中的研究文献。在 1969 年媒体接入控制问题首先出现在分组交换系统中之后，用并行信道通信的概念于 1983 年由 Marsan 和 Roffinela 在参考文献［1］论文《多信道局域网协议》中首次提出。媒体接入控制的传统问题阐述，正如第3章对媒体接入控制的阐述，是讨论多终端共享一个媒体。并非是使用所有的媒体，多信道媒体接入控制协议将原信道拆分为多个子信道并加以使用。对于同样的无线电资源，多信道媒体接入控制协议的性能被证明是优于单信道协议的。随着无线通信的发展，媒体接入控制有了更多新的范式。由于无线电传输范围是有限制的，一个节点（或者说是一个无线设备）并不一定能和其他节点通信。相比于有线网络，在无线网络中媒体接入控制的设计是一个更为复杂艰巨的任务。这是由于在无线传输媒体中，信号强度的衰减（路径损失，衰落）使得传输媒质具有高度的位置依赖性。因此，传统的传送前监听的机制（如 CSMA）不能很好工作，因为信道状态在接收端也许不同于其在发送端所估计的。这引出了所熟知的隐藏终端问题。

为了解决这个问题，无线局域网的随机信道接入协议通常使用在数据包发送前交换请求发送（RTS）和清除发送（CTS）控制报文的信道预留技术[18]。这个技术有效地执行了一种在接收端使得接收机进行集中式的媒体接入控制的虚拟载波监听[18]。此外，信道会暂时为数据包的传输预留，因为邻近的收到 CTS 的

接收机至少在数据传送期间会延误数据传送。这个期间可以包括在 RTS/CTS 报文中。IEEE 802.11 的无线局域网标准中，最根本的是使用了这个握手技术进行异步数据传输（即分布式协同功能 DCF)。

上述的关于单信道接入协议在重负载时接近无效率，碰撞将发生在控制数据报文中。随着可能不可同步的退避延迟，如果所有的竞争节点都退避传输，媒体可能完全空闲。更进一步，任何监听 RTS 和 CTS 的设备必须延迟到至少整个交换的结束（即 ACK 的结束)，这意味着当两个发送端互相监听时并发传输不可能发生，即使相应的接受机并没有监听它们的发射机之外的任何设备（称为暴露终端问题)。碰撞可能发生，并发传输很有限。多信道的使用提供了很多性能优势。多通道媒体接入控制协议减少碰撞，使得更多的并发传输，从而在相同的总容量中有更好的带宽利用率。多通道媒体接入控制协议允许很多节点在同一邻近地区不同的信道并发传输，而不互相干扰。载波监听加上一种有效的信道选择机制，可以挑选最干净的传输信道。多信道媒体接入控制协议现在开始一个新的发展阶段："如何选择信道?" 或者 "如何使所有设备都知道该怎么做?" 最近，所提出的多信道媒体接入控制研究更着重于如何使一个信道接入进行协商，就是说每个设备如何互相找到。请回顾一个事实，一个无线设备（或者节点）仅能够使用一个无线电频段，因为其半双工本质，在同时传输或接收时很容易受到攻击。随着软件无线电和认知无线电的发展，多信道媒体接入控制控制在认知无线电网络的媒体接入控制中起着关键的作用。

展示多信道媒体接入控制优点的两个重要的因素是：

● 单宽带信道与多信道的对比：一种直观的想法是，如果我们将一个固定带宽的无线电资源划分入很多子信道，并行传输是可能的。它的显著弊端在于，单个信道的容量/吞吐量被分解了。然而，整个系统的吞吐量是增加的。另外，在未来的通信网络中，设备能够接入宽带频谱中；并非使用一个可能受到频率选择性衰落的宽带频谱，该协议适用于具有多信道选择能力来选择良好的信道。多信道而不是单宽带信道直觉上是有道理的。

一个简单的例子可以证明这一优势。考虑如图 8-1 的情况。每个节点 A 和节点 B 在每个时隙具有概率 P 传送到节点 C；就是说，节点 A 和节点 B 服从独立同分布的二项式过程。对于 ALOHA 是无载波监听的。信道容量是 K，即对于每个成功的传输，吞吐量为 K 字节。延迟和重传可被忽视的。

总数据吞吐量为

$$2K(P)(1-P)=2KP-2KP^2$$

我们固定总信道容量为 K，将其拆分成两个相同的信道，每个信道容量为 $K/2$。我们同样使得节点 A 和节点 B 随机选择拆分后两个信道中的任意一个进行传输，如图 8-2 所示。假设节点 C 具备能够同时接收信道 1 和信道 2 信息的

能力。

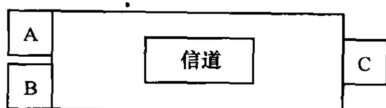


图 8-1 单通道随机媒体接入控制（时隙 ALOHA）

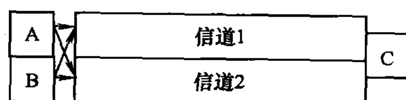


图 8-2 分通道随机媒体接入控制（时隙 ALOHA）

总数据吞吐量为

$$2\frac{K}{2}\left\{P\cdot\frac{1}{2}\underbrace{\left[(1-P)+P\cdot\frac{1}{2}\right]}_{\text{选择信道 \quad 成功}}+P\cdot\frac{1}{2}\underbrace{\left[(1-P)+P\cdot\frac{1}{2}\right]}_{\text{选择信道 \quad 成功}}\right\}$$

$$=2KP-KP^2$$

即使具有同样的聚集的信道容量，当概率 $P>0.667$ 时，多信道媒体接入控制协议明显优于单信道。在拥挤的网络情况下，即使概率 P 低于 0.667，多信道媒体接入控制协议仍然能优于单信道。更严格的结果显示在图 8-3 的一个 M-多信道、任意信道选择、CSMA 协议。从图中可以得出，对于固定信道，拆分信道越多，吞吐量越大。

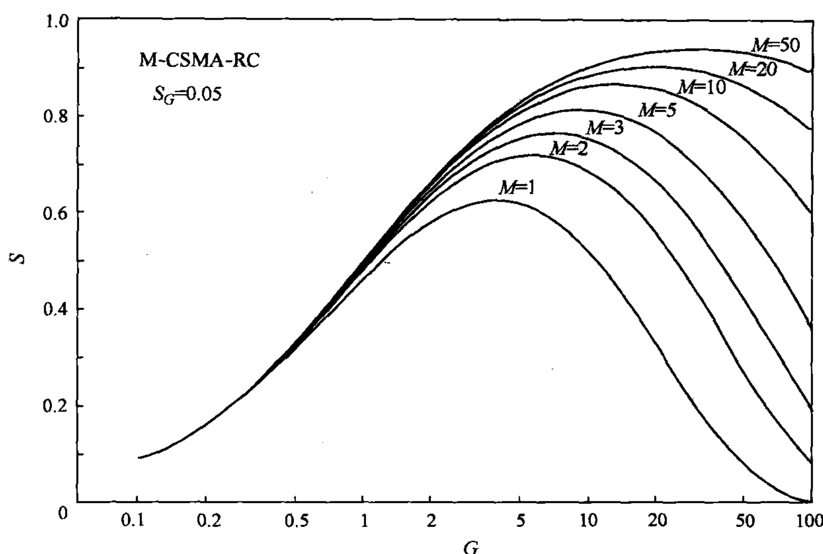


图 8-3 不同通道数的 M-信道多信道 CSMA 的吞吐量

● 划分现有的信道与多信道扩展的对比：下一个关注的问题：如何产生多信道。基于对多信道的使用方式，这些协议可大致分为两类。第一类假设在网络

中的每个设备有一个独立的信道，这是在单个的 CDMA 码（或正交信号）的基础上形成的。传输也许是传输机导向的（即每个设备使用自己的代码/信号传输）或者接收机导向的（即所有对于向同一接收器的传输使用相同的代码/信号）。在第一类中，所有信道都是专用的，也就是说，每个设备对于其代码有一个且只有一个唯一的信道。多信道媒体接入控制协议的第二类不假设每个节点的专用信道。相反，可用带宽假设被分为多个信道，信道的数量少于设备的数量。一个设备可以在任何信道中传输和接收。在移动 Ad hoc 网络中，对于每个设备，相对于专用信道使用少量的共享信道具有一些优点。举例来说，这并不需要每个设备知道在网络中使用的整套代码，而这将大大降低设计的复杂性。

8.2.1 多信道媒体接入控制的一般描述

为了不失一般性，我们从 K 个设备/节点的原始的媒体接入控制开始讨论，参见图 8-4。

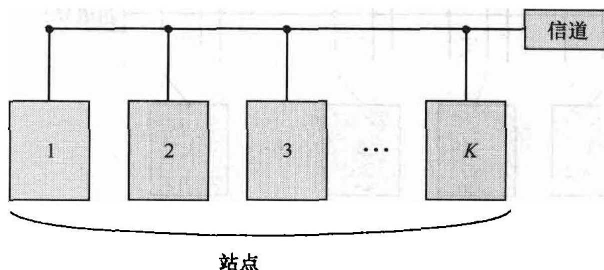


图 8-4　原始的媒体接入控制问题

所有的节点/设备共享相同的媒体，并且试图相互进行通信。如果超过两个节点试图接入共享媒体，碰撞将会发生。换句话说，就是存在有限资源竞争。目前，有多种协议（如 CSMA/CD 和 CSMA/CA）、多种技术（如载波检测、帧内间隔（IFS）和随机退避等）都试图解决此问题。与传统媒体接入控制不同的是，多信道方案提供了信道的多样性排序。假设信道数为 M，设备/基站数为 K。每个节点具有在多信道上感知、传输和接收的能力。如果存在任何共同信道的传输，碰撞仍将发生。这个多信道的情况如图 8-5 所示。

在无线网络中，图 8-5 中的情形并不合理，因为节点在传输时不能监听。此外，无线电数量是有限的，多数情况下只有一个。因此，每一个节点都必须调整到一个特定的信道中执行传输或接收。如果在同一信道中有多于两个传输，碰撞将会发生。图 8-6 描述了在无线（无线电）网络中这样的多信道的情况。

成功传输一个数据包，必须符合以下条件：

1）传输机可以找到接收机（接入协商）；

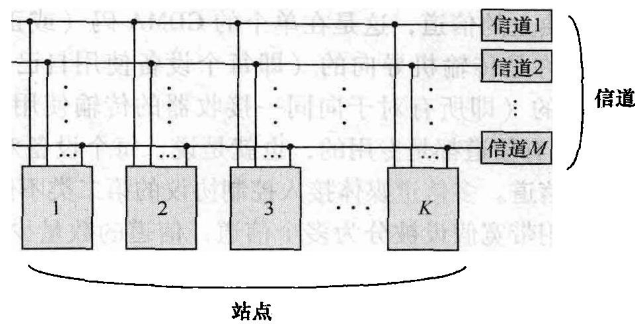


图 8-5 在局域网中的多信道媒体接入控制

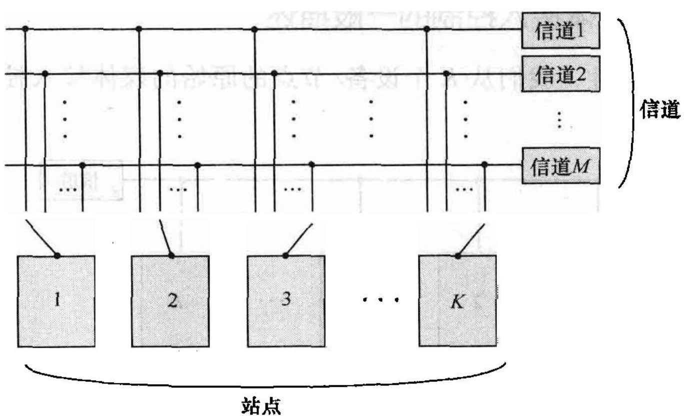


图 8-6 一般的多信道媒体接入控制（无线网络）

2）同一信道上没有同时传输（碰撞避免/解决）；

3）接收机必须与传输机在同一信道上（接入协商）。

一个成功的传输，如图 8-7 所示，碰撞或者错误通常通过 ARQ 处理。

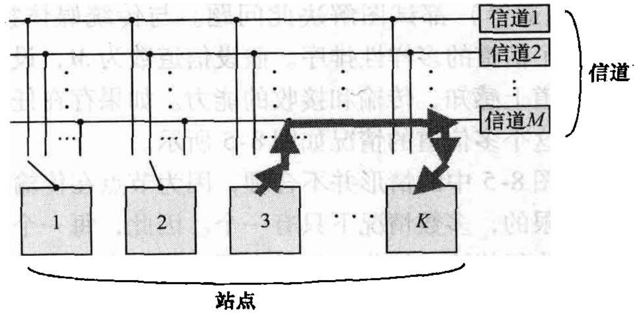


图 8-7 在信道 M 上的节点 3 到节点 K 的一个成功传输

上面的列出的第二个条件（碰撞避免/解决）是继承自原来的媒体接入控制问题，在 6.3.2 节中所述。当多于两个节点尝试接入信道碰撞必须避免/解决。多信道媒体接入控制协议第一部分是碰撞避免/解决，这还是属于 Gallager 发表在 1985 年 IEEE 学报的信息理论论文中的开创性观点的范畴[36]。

第一个和第三个条件（接入协商）中，对于有限的信道接入能力，一个节点可以通过一个或几个选定的信道监听/发送，但是为了接入网络，一个规划/协议必须得到妥善的设计。也就是说，发射机和接收机必须有一个机制来互相查找，这就是众所周知的多信道媒体接入控制的接入协商。因此，多信道媒体接入控制包括：

- 碰撞避免和碰撞解决；
- 接入协商。

我们可以注意到，在图 8-8 所示的无线多信道媒体接入控制的结构相当于一个分组交换网络和电路交换网络的混合体。

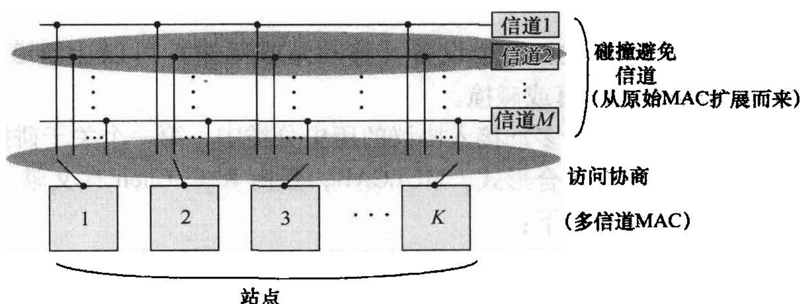


图 8-8 多信道媒体接入控制碰撞避免/解决和无线网络接入协商的两个基本部分

虽然多信道媒体接入控制协议中的碰撞避免/解决是继承自原始媒体接入控制问题，但它有以下几个方面的不同：

- 在一个信道上相互竞争的节点数量不一定是固定的，即发射机也许离开当前争用的信道，并尝试在下一帧中在其他信道接入它的接收机。
- 信道可能是不对称的，有不同的容量、时延、抖动等。因此，一个自适应结构的媒体接入控制协议可能会被要求增加对不同信道控制的效率。

在多信道中接入协商肯定是一个新的无线多信道问题，其中一个节点只能在一个信道监听/发送。该协议必须保证传输/接收的可接入性。换句话说，节点必须能够互相找到。

正如图 8-8 所示，广义的多信道方案是由一组参数 $\{M, q, C, N\}$ 确定；其中 M 和 N 分别表示信道的数量和设备的数量，$\{q_k \mid k=1, 2, \cdots, M\}$ 和 $\{C_k \mid k=1, 2, \cdots, M\}$ 表示其他干扰出现的概率和信道 1，2，…，M 的归一

化容量。这些参数可以完全确定多信道环境。干扰 $\{q\}$ 表示其他系统（例如在认知无线电网络中的主系统）出现的概率。

8.2.2 多信道媒体接入控制：碰撞避免/解决

这部分的多信道媒体接入控制相同于在第 3 章与 6.3.2 节中描述的传统的媒体接入控制（或多址接入）。

8.2.2.1 多层碰撞避免/解决 （MULCAR）

基本的媒体接入控制设计目标是在共享媒体中通过预期碰撞避免或者碰撞解决来减少碰撞的影响。这正是在 Gallager 的论文中建议的载波监听和碰撞解决。在 CSMA 的族中，传输前监听用于避免预期的碰撞。这个碰撞解决背后的想法是，有效地解决碰撞，进而通过树扩展结构的分裂算法提高多址接入效率。数学上，媒体接入控制协议的设计有以下信息：

■ 载波监听的一般形式：一个设备可以感知信道是“0”或“t”，即信道上是否存在传输。

■ 碰撞解决的一般形式：一个设备（移动站或基站）可以让信道为“0”、“1”或“e”，即空闲、占有或碰撞。

通过树展开结构起源于多址接入协议的历史分类中，第一个关于碰撞避免和解决的树展开结构的一般组合形式，MULCAR，是由 K. C. Chen 在文献［22］中提出的，并且进一步阐述如下：

■ 碰撞预测树展开（CATE）：此树结构避免/减少可能的碰撞，这主要是禁止信道效率的原因。这种树的扩展提供很多子集（叶子）在传输识别阶段之前分开碰撞的用户。

■ 碰撞解决树展开（CRTE）：此树结构在 20 世纪 70～80 年代得到很好的研究。不同于 CATE，CRTE 只在传输或识别阶段或一些碰撞被探测到之后生成树。设备等待信道反应，尝试像在第 3 章中阐述的 CRP 一样解决碰撞。应当注意到，分裂树可能不是一个二叉树。

■ 多层碰撞避免/解决（MULCAR）：考虑到 CATA 和 CRTE 的一般组合的形式，讲述 CATE 和 CRTE 的，多层碰撞避免/解决的总体结构描绘如图 8-9 所示。这个想法很简单：创建几个独立的“瓮”（组），然后尽量避免/解决竞争。MULCAR 可能包括有效扩展的以下几层：

—频域扩展；

—时域扩展；

—空域扩展；

—通信信道；

—随机退避概率域；

一信号域。

事实上，多信道媒体接入控制问题是 MULCAR 的一个特例。主要的区别在于分裂算法：如何把设备分成不同的群体以避免/解决竞争。这是“接入协商”的另一个观点。已经存在一些解决方案，例如在集中式网络中的随机寻址轮询[18,23]。但在分布式网络中，该问题尚未解决。

8.2.2.2　基于 CSMA 的随机选择多信道媒体接入控制

请回顾图 8-5 中描述的多信道媒体接入控制情况，它包含一组 M 个并行的信道向 N 个节点广播，其中每个节点都有独立的接口（对应每个信道）。这些广播信道拥有不同的带宽，设第 i 个信道的带宽为 W_i。那么，总的可用带宽为

$$W = \sum_{i=1}^{M} W_i$$

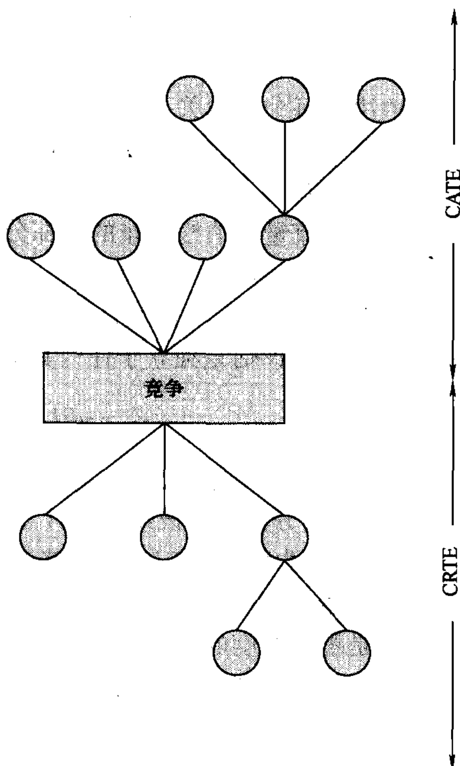


图 8-9　多层碰撞避免/解决（NULCAR）

假设数据包有恒定长度为 b 比特，那么在第 i 个信道上的数据包传输时间 T_i正比于 $1/W_i$。如果所有的信道拥有相同的带宽，即 W/i，$i=1, 2, \cdots, M$，T_0为拥有带宽 W 的单广播信道上传输一个数据包的时间，那么 $T=MT_0$。假设信道端到端的传输延时为 D 秒。归一化的传输延时为 $d_0=D_0/T_0$和 $a_i=D/T_i$。

假设一个无限的数量模型，新的数据包以泊松过程 λ 包/秒的速度产生。在均衡条件下，系统总吞吐量以 T_0单位归一化时间，被设定为 $S=\lambda T_0$。定义 λ_i为在第 i 个信道上稳定地被成功传输的数据包数量。第 i 个信道上总吞吐量为 $S_i=\lambda_i T$，还有，

$$W_i = \frac{W}{M}; T_i = MT_0; d_i = \frac{d_0}{M} \tag{8-4}$$

$$\lambda = \sum_{i=1}^{M} \lambda_i \tag{8-5}$$

进一步得到：

$$S = T_0 \sum_{i=1}^{M} \frac{S_i}{T_i} \tag{8-6}$$

在给定的广播系统中总数据流量，包括新的和重新调度的数据包，是假定为速率 γ 包/秒的泊松分布。因此，所提供的流量是 $G=\gamma T_0$。每个信道上流量的随机特性取决于每个基站关于选择一个信道并在其中调度数据包传输的选择策略。如果信道是随机选择的，独立与它们的状态，在第 i 个信道的流量是速率为 γP_i 的泊松分布，其中 P_i是选择信道 i 的概率。如果一个信道是随机选择，那么每个信道所提供的流量是泊松分布的。在分析中这个对称信道的属性是很重要的。如果信道不是相同或对称的泊松分布，这种分析就变得多余。

如果假设在信道 i 上流量是呈泊松分布的，对 S_i 的评估将会大大简化。在这种情况下，S_i 的评估是独立于任何其他信道，采用单信道的流量是

$$G_i=\gamma PL_iT_i=\frac{GP_iT_i}{T_0}=\frac{GP_iW}{W_i} \tag{8-7}$$

至于与其他参数相关的信道 i（例如传输延时归一化为 T_i: $d_i=d_0T_0/T_i$，等）。这个泊松分布的假设用于估计当每个信道的流量不是泊松分布时的性能。

8.2.2.3 非坚持 CSMA 随机信道选择的多信道媒体接入控制协议

在单信道非坚持 CSMA 协议中，当一个节点准备进行数据包传送时，无论是新的或者是重新调度的，它首先监听信道。如果监听得知信道空闲，数据包将立即传输。如果监听到载波（信道正在被其他基站使用），数据包将在随机的一段延时之后调度重传。两节点间由于非零的传输延时数据包可能碰撞。碰撞的数据包将会在一段随机等待之后重传。多信道系统意味着一个节点必须选择一个信道传送数据包。有多种信道选择的方法，这里我们只介绍一种随机信道选择的最简单情况。

在多信道 CSMA-RC（随机信道选择）中，一个有待发的数据包的节点在发送前随机地选择一个信道；此过程称之为随机信道选择（RC），随之产生的协议被称为 M-SCMA-RC。信道 i 以概率 P_i被随机选择；第 i 个信道的流量是速率 G_i 的泊松分布，如式（8-7）中的每一个数据包每一时间 T_i。信道 i 的吞吐量是从类似于原始 CSMA 的非坚持 CSMA 单信道中，通过用 G_i替代 G、用 a_i替代 a 后，推导出的。总吞吐量 S 表示为

$$S=G\sum_{i=1}^{M}\frac{P_i e^{-\frac{a_iGPW_i}{W_i}}}{GP_i(2a_i+1)\dfrac{W}{W_i}+e^{-\frac{a_iGP_iW}{W_i}}} \tag{8-8}$$

在两个信道的情况下（$M=2$），通过取 S 的对 P_l的偏微分为零可得到最大化吞吐量时的概率 P_1和 $P_2(P_2=1-P_l)$。图 8-10 所示为 P_1最优解。我们可以从结果中观察到的最佳信道的选择大约等于信道带宽的一个比例，当其所在的流量 G 低时。当部分信道带宽赋予等值时，即 $W_1=W_2=W/2$，最优的概率为 $P_1=P_2=0.5$。

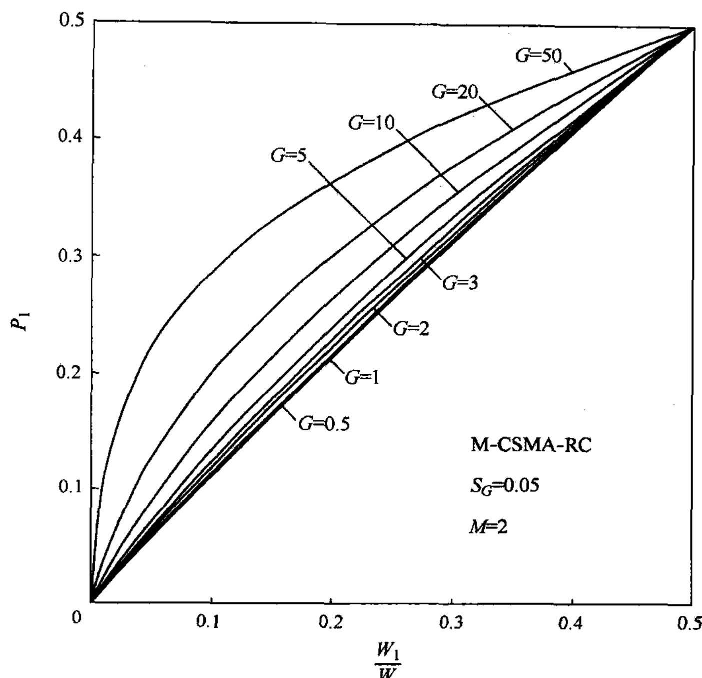


图 8-10　选择通道 1 的最优概率最大化总吞吐量

可以进一步证明相等的概率可在 M 个相同分解的信道里实现最大化。在两个信道的情况，即 $P_1=P_2=0.5$。

当 P_i的值为 $1/M$ 时，最优吞吐量可以简化为

$$S=\frac{Ge^{-\frac{a_0G}{M}}}{G(2\frac{a_0}{M}+1)+e^{-\frac{a_0G}{M}}} \tag{8-9}$$

平均数据包延时分析如下。我们观察到，每一次数据包调度传输并且成功传送的概率为 S/G。调度中传输的数据包的数量是一个均值为 G/S 的几何随机分布。当信道监听时，设选定的信道为空闲的概率为 P_{li}。那么，P_{li}表示如下：

$$P_{li}=\frac{1+a_iG_i}{G_i(2a_i+1)+e^{-a_iG_i}} \tag{8-10}$$

调度导致无传输的均值是 $G(1-P_{li})/S$。一个数据包发生碰撞的平均值是 $GP_{li}/S-1$，其中 1 表示最后一次成功的传输。平均数据包延时归一化为 T_0 的表达式如下：

$$D_0=\underbrace{\underbrace{\left(\frac{G}{S}P_{li}-1\right)}_{\text{碰撞}}(M+X_0+\tau_0+a_0+a_0)}_{\text{重传的时间请求}}+\underbrace{\underbrace{\left(\frac{G}{S}(1-P_{li})\right)}_{\text{无传输}}X_0+M+a_0}_{\text{一次成功传输的时间请求}} \tag{8-11}$$

其中，X_0是归一化为T_0的重传延时平均值，τ_0是请求ACK传输时间并归一化到T_0，P_{li}是基站监听被选的信道是空闲的概率。

根据M-CSMA-RC将一个单信道CSMA拆分成M个并行接入链路，增加了吞吐量并减少了每个信道的归一化的传输延时。结论显示，平均归一化的数据包延时也可以提高，只要操作吞吐量程度不是太低。

8.2.3 多信道媒体接入控制：接入协商

多信道媒体接入控制协议的第二部分是接入协商，即节点如何履行正确的接入协商，这相当于节点如何找到对方。首先，我们引入一个新的多信道媒体接入控制中的隐藏终端问题，该问题必须在设计多信道媒体接入控制时加以解决。然后，我们将介绍一些最近的多信道媒体接入控制协议。

8.2.3.1 多信道隐藏终端问题

我们描述一种新的典型的涉及多信道环境的隐藏终端问题。为方便说明，我们从一个简单的但不能解决这个问题的多信道媒体接入控制协议开始。

该协议类似于媒体接入控制802.11中使用一个收发机的DCF。假设有N个信道可用，其中一个专用于控制信息交换（控制信道），而其他信道则用于数据通信。当一个节点既不传输又不接收时，该节点监听控制信道。当节点A想要传送一个数据包给节点B，类似于IEEE802.11中的DCF，节点A和节点B相互交换RTS和CTS以预留信道。RTS和CTS消息被送到控制信道。在发送RTS时，节点A包括了一个想要使用的信道列表。在收到RTS后，节点B会选择一个信道并包括所选的信道信息于CTS中。此后，节点A和节点B切换它们的信道到一致认可的数据信道，然后交换（DATA）和ACK数据包。握手一旦完成，节点A和节点B立即切换到控制信道。

如图8-11所示，节点A有一数据包待送至节点B。节点A在信道1上，即控制信道，发送一个RTS。节点B选择信道2用于数据传送，并回送一个CTS给节点A。RTS和CTS的消息预留了信道2给节点A和B进行数据传输；这时无碰撞发生。然而，如果当节点B回送CTS给节点A时，节点C忙于在另外一个信道接收数据，所以它没有监听到CTS。在不知节点B占用信道2的情况下，节点C可能已经开始与节点D协商通信，并且选择信道2作为通信信道。这样就导致了在节点B发生碰撞。

上述问题发生是由于节点可能监听不同的信道使得虚拟载波监听难以实现，而无法避免隐藏终端问题的发生。如果每个节点只监听一个信道，那么节点C将可以监听到CTS，但同时也削弱了它的传输能力。这就是我们称以上问题为多信道隐藏终端问题的原因。

多信道隐藏终端问题必须在设计多信道媒体接入控制协议的接入协商部分时

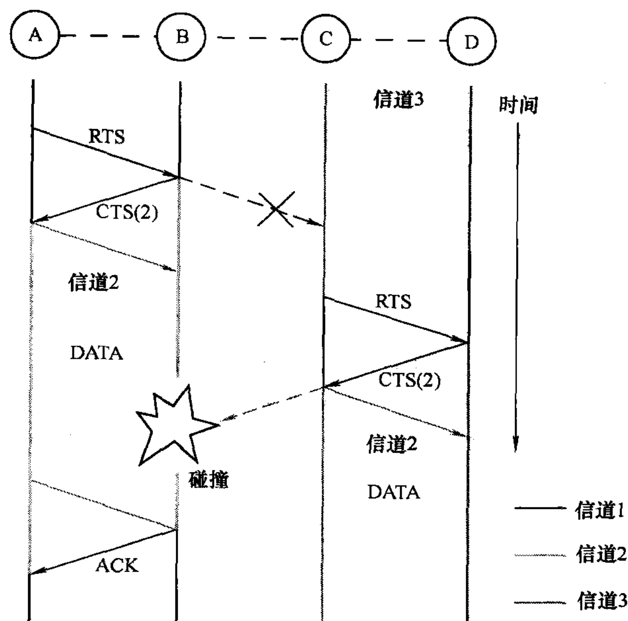


图 8-11　多信道隐藏终端问题

加以解决。在分布式网络中，如多信道媒体接入控制的 Adhoc 模式，该问题将变得严重。因此提出了许多不同的多信道协议用于缓解多信道隐藏终端问题。

8.2.3.2　接入协商机制

多信道媒体接入控制协议中的接入协商机制分为两类：窄带方法和宽带方法。在使用窄带发射机和接收机的情况下，用于传输的频带可以被预先确定或动态地选择。而在宽带系统中，传输端会在探测到未被占用的多个频带上进行数据传输。注意在宽带系统中，每个设备可能需要多于两个无线电波段以实现多频段/信道上的并行传输。而在窄带系统中，也许会要求少于两个的无线电波段。

8.2.3.3　窄带接入协商机制

专用控制信道

Wu 等在参考文献［7］中提出一种协议，可以根据需求动态地分配信道。这个协议被称为动态信道分配（DCA），它保持一个专用的控制信道，其他信道用于数据通信。每个主机有两个接收机，所以能够同时监听控制信道和数据信道。RTS/CTS 包可在控制信道上交换，数据包在数据信道上传送。在 RTS 包中，发送端包括一个首选信道的列表。当接收到 RTS 包后，接收端选取一条信道，并包括了信道信息于 CTS 包中。从而，数据（DATA）和应答（ACK）包在协定的信道上进行交换。由于两个接收机中总有一个在监听控制信道，多通道隐藏终端问题并不会发生。图 8-12 所示为专用控制信道协议的原理。

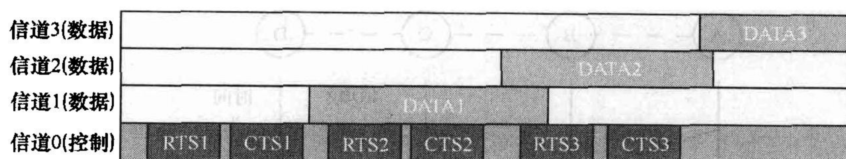


图 8-12　专用的控制通道

该协议不需要同步，而且在仅用少量控制信息的情况下就能使用多信道通信。但是，该协议在所有信道有相同带宽的情况下性能并不太好。当信道数量不多时，选用一个信道作专用控制信道就显得很浪费。在 IEEE 802.11b 情况中，仅有三个可用信道，那么选用其中一个作为控制信道就意味着占用了总带宽开销的 1/3。另一方面，如果信道数量很多，控制信道会成为阻碍数据信道充分利用的瓶颈。

Jain 等在文献［6］中提出了一个类似的协议，具有一个控制信道和 N 个数据信道。不同的是，根据接收端的信道情况选择最佳信道。这个协议智能地选择数据信道，有力地提升了数据吞吐量，但还是存在如动态接入控制（DCA）同样的弊端。

共同跳频

跳频协议的原理（见图 8-13）在于利用跳频自身特性将流量分配在多信道/频带上。跳隙预留多址接入是一个网络中用于低速跳频扩频（FHSS）的多信道协议。根据既定的跳频模式，主机从一个信道跳到另一个信道。当两个主机同意交换 RTS/CTS 握手后，它们停留在一个跳频上进行数据传输。其他主机继续跳频，多个通信可以在不同的跳频上进行。接收机用双轮询方式初始化信道跳频，除了是接收机而不是发射机来初始化碰撞避免的握手协议。这个方案能够使每个主机仅用一个接收机实现，但是仅适用于跳频网络，不能用于其他机制，如直接序列扩频（DSSS）通信。

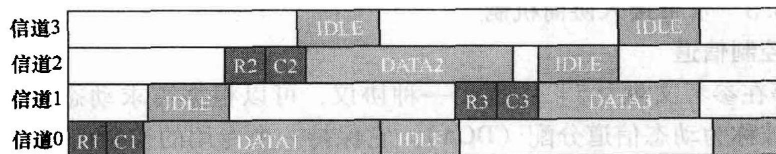


图 8-13　共同跳频协议

分阶段

在这个方法中，设备使用一个无线电频段。时间被划分为一个交互的控制和数据交换的阶段序列。在控制阶段中，所有设备调制到控制信道，尝试使协商用于接下来的数据交换阶段的信道。如果设备 A 有一些数据要发送给 B，它首先通

过最低数字编号的空闲控制信道，假设信道 i，发送一个数据包到设备 B。设备 B 回送一个确认数据包给 A。如此，A 和 B 达成共识，将在接下来的数据交换阶段中使用信道 i。一旦确定后，设备就不能接受与之冲突的其他协议。（注意，当存在隐藏节点时，发送端和接收端可能在哪些信道是空闲的判断上有很大分歧。那么，这就需要一个更为复杂的协议，如 MMAC。）图 8-14 描述了分阶段的原理。

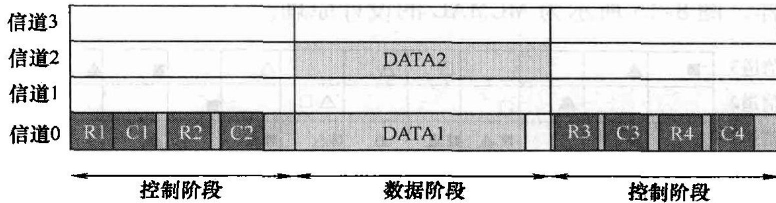


图 8-14　分阶段协议

在第二阶段，设备都调制到协定的信道开始数据传输。协议允许多对设备选择同一信道，因为每一对也许没有足够多的数据用尽整个的数据阶段。结果是，不同的通信对必须在数据阶段要么自己协调，要么竞争。这个方法的优点在于每个设备仅要求一个无线电波段。然而，这需要时间去同步所有设备，尽管同步比在共同跳频中要松散一些因为设备跳频不太频繁。这个方法的例子是 MMAC 和多信道接入协议（MAP）。它们的主要区别是在 MMAC 中数据传输阶段的周期是固定的，然而在 MAP 中这个周期是根据在控制阶段中的协定而变化的。

混合协议

混合协议考虑了多信道多波段的分配和在一个多层方法中的接口的分配。假设路由信息是众所周知的，每个节点可以动态地建立信道分配表，并通知整个网络。在媒体接入控制层，这个方法可以动态地或者以现有的多信道协议的混合模式来选用。注意，混合协议中的稳定路由信息假设前提是无线网状网络，这并不是普遍认为的更一般化的 Ad hoc 模式。

SSCH

有许多在每个节点上遵从信道数的跳频序列。每个序列是基于一个伪随机发生器种子而唯一决定的。每个节点选中多个序列，然后以一种时间复用的方式遵照序列。当节点 A 想要和节点 B 通信，A 等待直到 B 处在同样的信道上。如果 A 频繁地想和 B 通信，A 采用一个或多个 B 的序列，因此它们花在同一信道上的时间增加了。为了让这个机制工作，发射机需要通过种子广播的机制学习接收机的当前序列。

MCMAC

每个节点选中一个种子产生不同的伪随机跳频序列。当一个节点空闲时，它

遵从原先跳频序列。节点在每个发送的包中包括了它的种子，所以它的近邻最终将了解它的跳频序列。然而，节点并不需要对齐它们在实际中的跳频界限。当节点 A 有数据传送给节点 B 时，A 随机地在每个时隙以概率 P 传输数据。如果决定传输，它将调制到 B 的当前信道，发出 RTS 消息。如果节点 B 回应了一个 CTS，节点 A 和 B 停止跳频而交换数据。数据的交换一般发生在多个时隙中。数据交换结束后，节点 A 和 B 返回到它们的原始跳频序列，就好像跳频中没发生过中断一样。图 8-15 所示为 MCMAC 的设计原理。

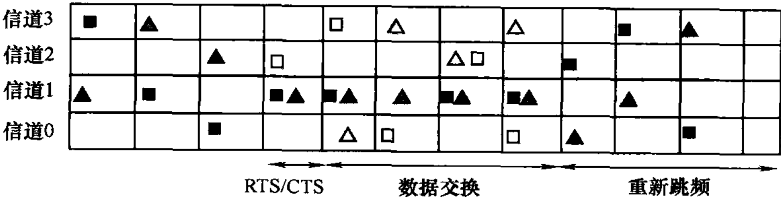


图 8-15　多信道媒体接入控制

SSCH 和 MCMAC 是相似的，因为它们都允许节点在不同的信道交汇。然而，这之间有微妙的差别。在 SSCH 中，每个节点选择四个不同的跳频序列，然后以时间复用的方式形成单个跳频序列。节点在时间上采用它们的跳频序列进行交流，但是它们不被允许偏离跳频序列。在 MCMAC 中，每个节点有一个从不变化的跳频序列。然而，节点是被允许暂时偏离跳频序列，从而方便传输和接收数据。在 SSCH 中，一个发射机在传输前必须等待直到它的接收机也在信道上。在 MCMAC 中，发射机能够暂时的偏离它的序列而跳到它的接收机的信道上发送数据。

认知多信道媒体接入控制

根据认知无线电的原理，节点应该能够监测环境并能相应的重新配置它的结构。认知多信道媒体接入控制是作为 MCMAC 的升级版出现的。每个协议的运行框架被分为三个阶段：频谱感知（SS）阶段、CSMA 竞争阶段和数据传输阶段。在认知多信道媒体接入控制中，设备能够感知外部环境、网络的信息和其他干扰源信息（例如主系统）。在收集并交换完如上信息后，节点能够把传输概率设定为系统的最大吞吐量。图 8-16 所示为认知多信道媒体接入控制的设计原理。

8.2.3.4　宽带接入协商机制

CDMA

在多信道系统中采用 CDMA 的思路很简单。这些协议不使用分隔的信道，而是使用像多信道一样的 CDMA 正交特性。通过不同的扩频码，节点能够以不同的速率传输，即较短的编码表示较大的带宽，正如宽带 CDMA 蜂窝通信系统一样。然而，这个方法的主要缺点是它几乎不可能有全局码字信息，而且效率低下。

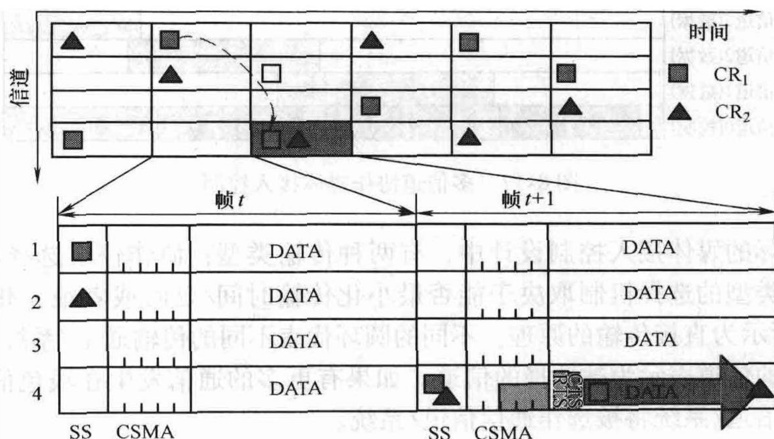


图 8-16　认知多信道媒体接入控制

M-信道选择

Nasipuri 等提出了一种“软”信道预留的多信道 CSMA 协议。如果有 N 个信道，协议假设每个主机能够同时监听到所有 N 个通道。一个想要传送数据包的主机搜索一个空闲的信道，然后在那个空闲的信道上传输。在空闲的信道中，一个被使用的完成了最后一次成功传输的信道是首选。这个协议能够被扩展至基于发射端观察的信号功率选择最佳的信道。这些协议使得每个主机需要 N 个收发机，这也许值得关注。

频率编码

这是 M-通道选择协议的升级版本。在一个宽带系统中，传输机可在探测到非占有的多频带上传输，即信号/码字在多频率上发送。接收机监控所有监测到的空闲频带。请注意，这样一个频率编码方案是不同于跳频的，它需要每次传输前信道在所有不同的频带都可用。N-收发机的主要缺点也是硬件需求，这正如 M-信道选择的缺点一样。

频率协作（协作式多信道媒体接入控制）

这是一种新的融合多信道和协作通信两方面考虑的多信道媒体接入控制协议。多信道中的一个开放性问题是非对称特性：有些信道更利于传输。然而，另外一个问题在于路由：如果目的地不在邻近该怎么办？其实，这两个问题是一致的：就是决定从哪个信道/设备传输。协作通信/网络已被证实能通过重复/回复来自传输机的信息，来提高系统吞吐量。一般的媒体接入控制设计遵从专用控制信道的方案，如图 8-17 所示，其中假设每个设备有无线电波段/接口。值得注意的是，在不同信道上传输的速率是不同的。由于每个节点能连续监听同一控制信道，多信道隐藏终端问题在此得以解决。

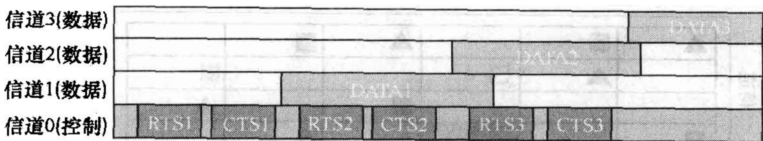


图 8-17 多信道协作媒体接入控制

在实际的媒体接入控制设计中，有两种传输类型：协作信道选择和直接传输。不同类型的选择机制取决于能否最小化传输时间/延时或者最大化吞吐量。图 8-18 所示为直接传输的原理。不同的圆环代表不同的传输通道/系统。在本例中，浅色的信道表示为被选择的信道。如果有更多的通信发生在浅色信道/系统内，深色信道/系统将被选作通信信道/系统。

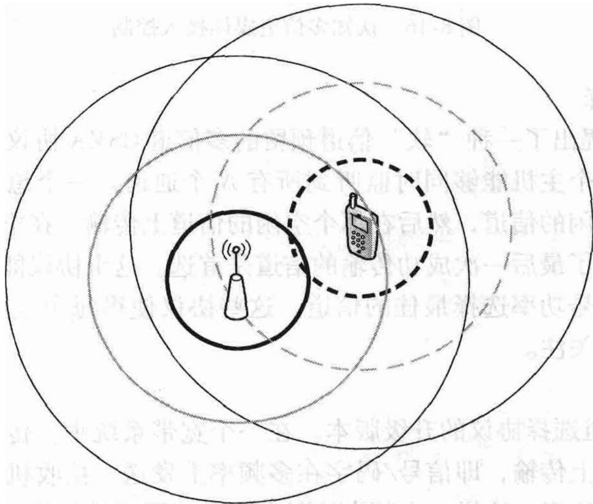


图 8-18 协作多信道媒体接入控制中的直接传输

图 8-19 和图 8-20 所示为协作多信道媒体接入控制中的协作传输。传输端可以计算期望的传输时间。如果在高传输率信道/系统中存在一个协作节点，它将发送求助请求。如果一个协作节点答复可以协作，传输机将发送数据给协作节点，然后协作节点将帮助数据中继。这些节点的位置/状态将导致不同的协作。例如如图 8-19 所示，节省总时间，如图 8-20 所示，传输端可以访问不在它传输范围内的节点。

8.2.3.5 分析和比较

本节我们将描述和分析简化的多信道媒体接入控制协议的模型。我们的目标是量化比较在同样的条件下，不同的多信道协议的性能。研究的几个假设条件如下：

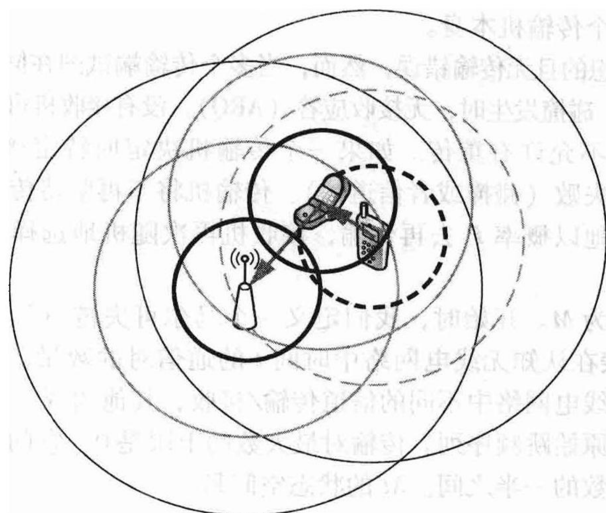


图 8-19　协作多信道媒体接入控制（节省总时间）

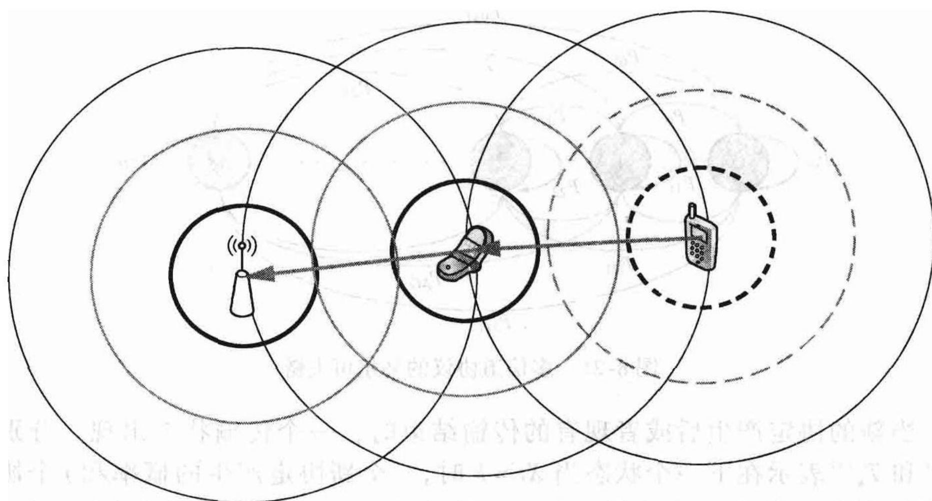


图 8-20　协作多信道媒体接入控制（传输接入节点不在它的传输范围）

■ 它是一个完全的分布式网络，其中所有的设备可以相互传输和接收。每个设备包含单个无线电波段。它能够在一个时隙接入到一个特定的可用信道。简单地说，我们假设网络是完全的分布式的且没有隐藏/暴露的终端。媒体接入控制因此简化成了一跳的问题。

■ 在网络中是完全同步的，而且划分为帧/时隙。

■ 设备一直有一个数据包要传输到其他设备，但是在每个时隙的开始时它们以概率 P 尝试传输。传输机的接收端是被传输机等概率的在全部设备中随机

地选择，除了这个传输机本身。

■ 信道是理想的且无传输错误，然而，当多个传输端试图在同一信道上同时传输时会引发碰撞。碰撞发生时，无接收应答（ARQ），没有接收机可以收到数据包。

■ 超出时隙不允许有重传。如果一个传输机决定向特定接收端发送信息，某种情况下传输失败（碰撞或者信道忙）。传输机将不再坚持传输去修正过去的失败。它会随机地以概率 P 去再传输，接收机再次随机地选择，而并不考虑之前的接收机。

可用信道数为 M。开始时，我们定义一个马尔可夫链 $\{Xt\}$，如图 8-21 所示，状态 Xt 代表在认知无线电网络中时间 t 的通信对的数量。当 $Xt=k$，有 $2k$ 个设备在认知无线电网络中不同的信道传输/接收，其他的 $N-2k$ 个设备保持沉默和遵从它们的原始跳频序列。传输对最大数的上限是 0，它的最小数量介于可用信道数和设备数的一半之间。Xt 的状态空间是

$$S\triangleq\left\{0,1,2,\cdots,\min\left(\left[\frac{N}{2}\right],M_D\right)\right\} \tag{8-12}$$

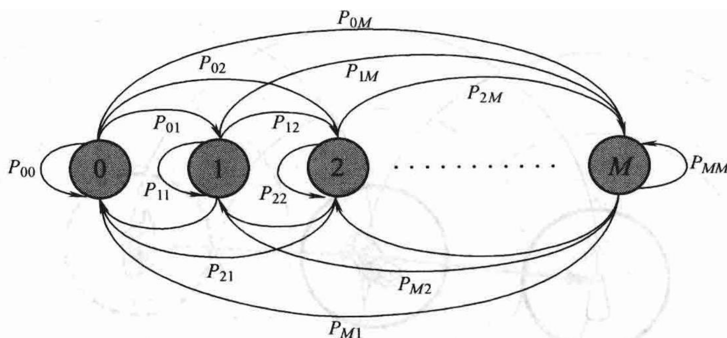


图 8-21 多信道协议的马尔可夫链

当新的协定产生后或者现有的传输结束时，一个传输状态出现。分别设 $S_k^{(i)}$ 和 $T_k^{(j)}$ 表示在下一个状态当 $Xt=k$ 时，i 个新协定产生的概率和 j 个协定结束的概率。那么，从时刻 t 状态 k 到时刻 $t+1$ 状态 l 的转移概率可以表示为

$$P_{kl}=\begin{cases}\sum_{m=0}^{k}S_k^{m+l-k}T_k^m, l\geqslant K\\ \sum_{m=k-l}^{k}S_k^{m+l-k}T_k^m, l<k\end{cases}$$
$$=\sum_{m=(k-l)^+}^{k}S_k^{m+l-k}T_k^m \tag{8-13}$$

式中，m 是在时间 t 到 $t+1$ 结束传输的传输数量，因此它的值是介于 $k-l$ 到 k 的。在这个表达式中，m 是结束传输的传输数量，它的值介于 $k-l$ 和 k 之间。至少有 $k-l$ 的传输必须结束，使得在下一个时隙有 l 对，k 是结束传输的最大数量。同样，当系统在状态 k 时，i 传输结束的概率 $T_k^{(j)}$ 为

$$T_k^{(j)} = Pr[j \text{ 传输在时间 } t \text{ 结束} \mid X_t = k]$$
$$= \binom{k}{j} q^j (1-q)^{k-j}$$

S^l是多信道媒体接入控制协议设计的关键。它描述了一个传输是如何初始化的。由于在认知无线电网络中 MCMAC 的马尔科夫链转移概率是可知的，我们可以依据马尔科夫链的平衡方程计算出每个状态的极限概率。平均的使用效率为

$$\rho = \frac{\sum i \in S^{i \cdot \pi_i}}{M}$$

式中，π_i是系统处于状态 i 的极限概率，S 是马尔科夫链的状态。因此我们可以通过将信道传输率和可用信道数乘以 ρ，进一步推导出总系统的吞吐量。系统的总吞吐量为

$$R_d = MC\rho_d$$

我们把标记列于表 8-1 中。

表 8-1　标记的总结

符　号	符号含义
N	设备数量
M	可用信道数量
$T_k^{(j)}$	下一个时隙中 j 传输结束的概率
$S_k^{(i)}$	下一个时隙 i 协定达成的概率
P_{kl}	从状态 k 到状态 i 的传输概率
P	为每一个认知无线电设备尝试传输的概率
q	主用户出现的概率
C	每一个信道的信道容量（bit/s）
ρ	[0，1] 之间的信道使用率

通过比较 S^l 和多信道媒体接入控制协议的传输概率，我们能推导和比较它们的性能。这里我们忽略复杂的结果，但是在表 8-2 中列出了四个协议：专用控制信道、共同跳频、分阶段、MCMAC 的代表性结果。

图 8-22 所示为在 802.11a/b 下不同的多信道协议的性能。我们能看出 MCMAC优于其他的协议，专用控制信道是次优的。值得注意的是，当信道较多时，专用控制信道优于 MCMAC。

表 8-2 MCMAC 协议的分析模型

协议	$S_k^{(i)}$ 和 P_{kl}
专用控制信道	$S_k^{(i)}=\begin{cases}(N=2k)p(1-p)^{(N-2k-1)} & i=1\\ 1-S_k^{(1)} & i=0\\ 0 & \text{其他}\end{cases}$ $p_{kl}=\begin{cases}0 & l>k+1\\ T_k^{(0)}S_k^{(1)} & l=k+1\\ T_k^{(k-l)}S_k^{(0)}+T^{(k-l+1)}S_k^{(1)} & 0<l\leqslant k\\ T_k^{(k)}S_k^{(0)} & l=0\end{cases}$
共同跳频	$S_k^{(i)}=\begin{cases}(N-2k)p(1-p)^{(N-2k-1)}\times\dfrac{N-2k-1}{N-1}\times\dfrac{M-k}{M} & i=1\\ 1-S_k^{(1)} & i=0\\ 0 & \text{其他}\end{cases}$ $p_{kl}=\begin{cases}0 & l>k+1\\ T_k^{(0)}S_k^{(1)} & l=k+1\\ T_k^{(k-l)}S_k^{(0)}+T_k^{(k-l+1)}S_k^{(1)} & 0<l\leqslant k\\ T_k^{(k)}S_k^{(0)} & l=0\end{cases}$
分阶段	详见参考文献 [35]
McMAC	$S_k^{(j)}=\sum_{0,i,a}P[A=a]\times P[O=o\mid A=a]$ $\times P[I=i\mid O=o,A=a]P[J=j\mid I=i,\ A=a,\ O=o]$ 详见参考文献 [35] $P_{kl}=\begin{cases}0 & l>k+1\\ T_k^{(0)}S_k^{(1)} & l=k+1\\ T_k^{(k-l)}S_k^{(0)}+T_k^{(k-l+1)}S_k^{(1)} & 0<l\leqslant k\\ T_k^{(k)}S_k^{(0)} & l=0\end{cases}$

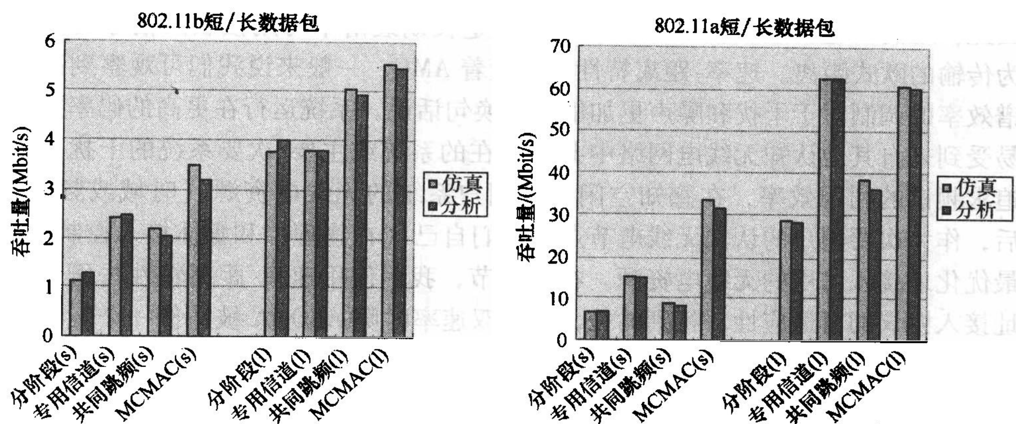


图 8-22　不同的多信道协议的性能

另外值得注意的是，MCMAC 对于多信道隐藏终端问题不是最佳的解决方案（见图 8-23）。一个可行的解决方案也许是专用控制信道，后者也能随着信道的增加而有增长。

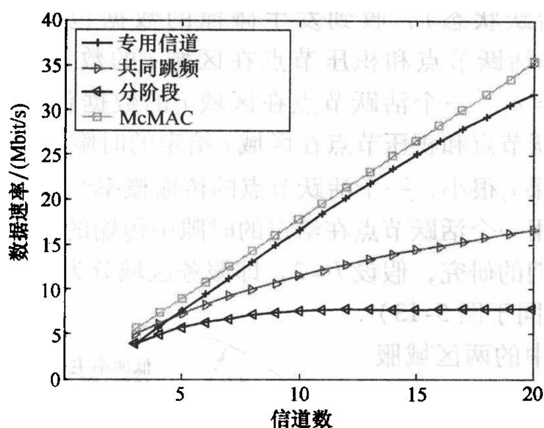


图 8-23　多个信道的不同多信道协议的性能

8.3　速率-距离自适应时隙 ALOHA

在无线通信网中，无线节点一般假设为在服务地区随机均匀分布的。因此，节点间的距离也是随机分布的。回顾在 5.4.3 节中的速率-距离特性，我们可以认为节点间的距离是接收信号的功率，而不是欧氏距离或者传输距离的一个度量。注意，传输距离被认为是传输机和接收机间具有同样接收功率的欧式距离。

因此，距离 D 意味着任何可能的位置在特定长期衰落中可用接收的信号功率作为传输的欧式距离。速率-距离特性常常随着 AMC，一般来说我们可观察到高频谱效率的调制对于干扰和噪声更加敏感。换句话说，系统运行在更高的码率更容易受到来自其他认知无线电网络中共同存在的系统或主要/次要系统的干扰。为追求频谱的利用效率，在感知空闲的主用户系统的无线电资源（时域或频域）后，作为次要用户的认知无线电节点将它们自己（在物理层和媒体接入控制层）最优化地接入可用的无线电资源。在这一节，我们使用速率-距离特性去研究多址接入协议的自适应性。特别的是，一个双速率时隙 ALOHA 被选作一个例子来演示多址接入协议的速率-距离特性。

8.3.1 系统模型

我们假设系统中节点的数量是 N。所有节点在服务区域随机地均匀地分布。服务区域分为 i 个区域。为简化起见，服务区域和每个区域假设是圆形的。每个区域 i 的半径是 d_i。一个时隙内新到达每个节点的数据包数量假设是一个泊松分布的变量。在区域 i 的节点数量是 n_i，有 $\sum n_i = N$。新近收到数据包的节点称为活跃节点（在活跃状态），收到发生碰撞的数据包的节点称为积压节点（处于积压状态）。活跃节点和积压节点在区域 i 的数量分别记为 $n_{a,i}$和 $n_{b,i}$。因此，有 $n_{a,i}+n_{b,i}=n_i$。一个活跃节点在区域 i 的数据包到达速率（或接入速率）为 r_i。一个活跃节点和积压节点在区域 i 给定的时隙中发送数据包的概率分别是 $q_{a,i}$和 $q_{b,i}$。如果 r_i很小，一个活跃节点的传输概率约为 $q_{a,i}=1-e^{-r_i}\approx r_i$。换句话说，我们能使用一个活跃节点在给定的时隙中传输的概率去表示数据包到达的速率。为简化我们的研究，假设 $I=2$，即服务区域分为两个区域。

如图 8-24（相同于图 5-13）所示为认知无线电中的两区域服务地区的速率-距离特性。主系统的基站和移动站通过主通信链路通信（PCL）。因为它们的有效距离，所以选择低速率。在基站的服务地区边界，有两个认知无线电设备在受到主系统的低干扰情况下正在建立一个次要通信链路（SCL）。如图 8-24 所示，高速率通信在两个认知无线电不影响主系统的情况下是可能的。

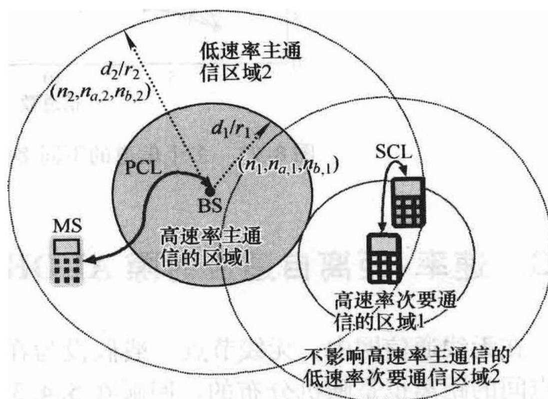


图 8-24 共存 PS 和 CR 系统的速率-距离特性

活跃主系统节点的干扰对于认知无线电是可容忍的。

为了分析速率-距离特性，如图 8-24 所示，一个二维马尔可夫链用来建模区域 1 和 2 中的积压节点的数量。当有 v 个积压节点，区域 i 中给定的时隙存在 u 个活跃节点在（$n_i - v$）个非积压节点传输数据包的概率是

$$P_{a,i}(u,v) = \binom{n_i - v}{u}(q_{a,i})^u(1 - q_{a,i})^{n_i - v - u} \tag{8-14}$$

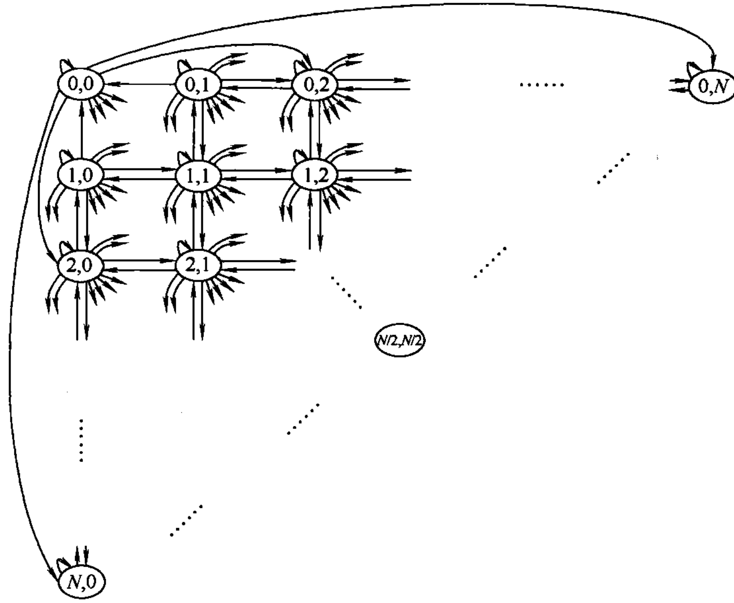


图 8-25　二维马尔可夫链模型

区域 i 中给定的时隙在 v 个积压节点中存在 u 个积压节点传输数据包的概率是

$$P_{b,i}(u,v) = \binom{v}{u}(q_{b,i})^u(1 - q_{b,i})^{v-u} \tag{8-15}$$

稳定状态的概率可用以下两个公式来计算：

$$P(n_{b,1}, n_{b,2}) = \sum_{i=0}^{n_1}\sum_{j=0}^{n_2}\left[P(i,j)P_{(n_{b,1},n_{b,2})}^{(i,j)} + P(n_{b,1}+1, n_{b,2})P_{(n_{b,1},n_{b,2})}^{(n_{b,1}+1,n_{b,2})} + P(n_{b,1}, n_{b,2}+1)P_{(n_{b,1},n_{b,2})}^{(n_{b,1},n_{b,2}+1)}\right] \tag{8-16}$$

$$\sum_{n_{b,1}=0}^{n_1}\sum_{n_{b,2}=0}^{n_2} P(n_{b,1}, n_{b,2}) = 1 \tag{8-17}$$

式中，$P_{(n_{b,1}+i,n_{b,2}+j)}^{(n_{b,1},n_{b,2})}$ 是从状态（$n_{b,1}$，$n_{b,2}$）到状态（$n_{b,1}+i$，$n_{b,2}+j$）的状态转移概率，推导如下。因此，当在区域 1 和 2 节点的数量是 n_1 和 n_2（或者 $N - n_1$）

时，积压节点的数量是 $n_{b,1}$ 和 $n_{b,2}$，成功转移概率为

$$P_{(n_{b,1}+i,n_{b,2}+j)}^{(n_{b,1},n_{b,2})}=$$
$$\begin{cases} P_{a,2}(0,n_{b,2})P_{a,1}(1,n_{b,1})P_{b,1}(0,n_{b,1})P_{b,2}(0,n_{b,2}) \\ +P_{a,2}(1,n_{b,2})P_{a,1}(0,n_{b,1})P_{b,1}(0,n_{b,1})P_{b,2}(0,n_{b,2}) \\ +P_{a,1}(0,n_{b,1})P_{a,2}(0,n_{b,2})[(1-P_{b,1}(1,n_{b,1})-P_{b,1}(0,n_{b,1})) \\ +(1-P_{b,2}(1,n_{b,2})-P_{b,2}(0,n_{b,2})) \\ +P_{b,1}(0,n_{b,1})P_{b,2}(0,n_{b,2})+P_{b,1}(1,n_{b,1})P_{b,2}(1,n_{b,2})], & i=0,j=0 \\ P_{a,1}(1,n_{b,1})P_{a,2}(0,n_{b,2})[(1-P_{b,1}(0,n_{b,1}))+ \\ P_{b,1}(0,n_{b,1})(1-P_{b,2}(0,n_{b,2}))], & i=1,j=0 \\ P_{a,1}(1,n_{b,1})P_{a,2}(1,n_{b,2})[(1-P_{b,2}(0,n_{b,2}))+ \\ P_{b,2}(0,n_{b,2})(1-P_{b,1}(0,n_{b,1}))], & i=0,j=1 \\ P_{a,1}(0,n_{b,1})P_{b,1}(1,n_{b,1})P_{a,2}(0,n_{b,2})P_{b,2}(0,n_{b,2}), & i=-1,j=0 \\ P_{a,1}(0,n_{b,1})P_{b,1}(0,n_{b,1})P_{a,2}(0,n_{b,2})P_{b,2}(1,n_{b,2}), & i=0,j=-1 \\ P_{a,1}(i,n_{b,1})P_{a,2}(J,n_{b,2}), & 2\leqslant i+j\leqslant n_1+ \\ & n_2-n_{b,1}-n_{b,1} \end{cases} \tag{8-18}$$

$$\begin{aligned} P_{\text{succ}}(n_{b,1},n_{b,2},n_1,n_2) &= P_{a,1}(1,n_{b,1})P_{b,1}(0,n_{b,1})P_{a,2}(0,n_{b,2})P_{b,2}(0,n_{b,2}) \\ &+P_{a,1}(0,n_{b,1})P_{b,1}(1,n_{b,1})P_{a,2}(0,n_{b,2})P_{b,2}(0,n_{b,2}) \\ &+P_{a,1}(0,n_{b,1})P_{b,1}(0,n_{b,1})P_{a,2}(1,n_{b,2})P_{b,2}(0,n_{b,2}) \\ &+P_{a,1}(0,n_{b,1})P_{b,1}(0,n_{b,1})P_{a,2}(0,n_{b,2})P_{b,2}(1,n_{b,2}) \end{aligned} \tag{8-19}$$

令尝试率 $\Lambda(n_{b,1},\ n_{b,2},\ n_1,\ n_2)$ 为在一个时隙当区域 1 和 2 中节点数量是 n_1 和 n_2（或 $N-n_1$），积压节点数量是 $n_{b,1}$ 和 $n_{b,2}$ 的尝试传输的数量。因此，

$$\begin{aligned} \Lambda(n_{b,1},n_{b,2},n_1,n_2) &= (n_1-n_{b,1})q_{a,1}+n_{b,1}q_{b,1} \\ &+(n_2-n_{b,2})q_{a,2}+n_{b,2}q_{b,2} \end{aligned} \tag{8-20}$$

当 x 很小时，$(1-x)^y\approx e^{-xy}$，式（8-19）近似为

$$P_{\text{succ}}(n_{b,1},n_{b,2},n_1,n_2)=\Lambda(n_{b,1},n_{b,2},n_1,n_2)e^{-\Lambda(n_{b,1},n_{b,2},n_1,n_2)} \tag{8-21}$$

这也可认为是两状态 ALOHA 的吞吐量。令 $\bar{n}_{b,i}$ 作为区域 i 的积压节点平均数量。平均尝试率 $\bar{\Lambda}$（$\bar{n}_{b,1}$，$\bar{n}_{b,2}$，n_1，n_2）为

$$\bar{\Lambda}(\bar{n}_{b,1},\bar{n}_{b,2},n_1,n_2)=(n_1-\bar{n}_{b,1})q_{a,1}+\bar{n}_{b,1}q_{b,1}+(n_2-\bar{n}_{b,2})q_{a,2}+\bar{n}_{b,2}q_{b,2} \tag{8-22}$$

式中，$\bar{n}_{b,i}=\sum_{n_{b,1}=0}^{n_1}\sum_{n_{b,2}=0}^{n_2}n_{b,i}P$（$n_{b,1}$，$n_{b,2}$）。

基于第 3 章的稳定性分析，为了让时隙 ALOHA 稳定，在式（8-22）中的平均尝试率必须小于或者等于 1。进一步地，当平均尝试率等于 1 时，时隙ALOHA的吞吐量（即 1/e）达到最大。

考虑当系统是稳定的，且处于稳态时的情况。假设所有积压节点有同样的传输概率 $q_{b,1}=q_{b,2}=q_b$。令 $k_r=q_{a,2}/q_{a,1}$ 和 $k_d=d_2/d_1$。为了将速率-距离时隙ALOHA应用到认知无线电网络中，我们令 $q_{a,2}\leqslant q_{a,1}$，$d_2>d_1$。在这个情况下，有 $0\leqslant k_r\leqslant 1$ 以及 $k_d\geqslant 1$。因此，式（8.22）可写为

$$\overline{\Lambda}(\bar{n}_{b,1},\bar{n}_{b,2},n_1,n_2)=[N(k_d^{-1})^2-Nk_d^{-1}k_r+k_r(N-\bar{n}_{b,2})-\bar{n}_{b,1}]q_{a,1}+(\bar{n}_{b,1}+\bar{n}_{b,2})q_b \tag{8-23}$$

令 $N=40$，$d_2=80$，$q_{a,1}=0.02$，$q_b=0.05$。首先，我们保持尝试率小于 1（即 $\overline{\Lambda}$（$\bar{n}_{b,1}$，$\bar{n}_{b,2}$，N_1，N_2）<1），因此系统是稳定运行的。之后，我们让尝试率接近 1（即 $\overline{\Lambda}$（$\bar{n}_{b,1}$，$\bar{n}_{b,2}$，N_1，N_2）$\rightarrow 1$），因此吞吐量能够最大化。速率-距离的关系如图 8-26 所示，其中可看出如果 k_d变小（即 $k_d<1.2$），数据包到达区域 1 和 2 到达率的差别就会加大。相反，如果 k_d变大（即 $k_d>3$），数据包到达区域 1 和 2 的到达率几乎是在同一水平。

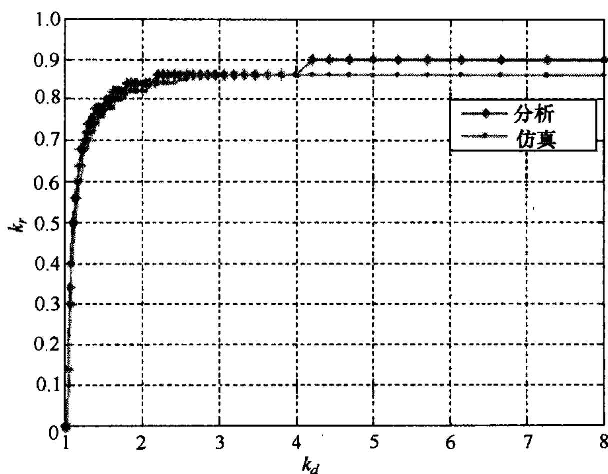


图 8-26　在稳定和最大化吞吐量限制下的速率-距离关系
（$N=40$，$q_{b,1}=q_{b,2}=q_b=0.05$，$q_{a,1}=0.02$，$d_2=80$）

延时分析在媒体接入控制协议性能中是很关键的。假设 $q_{b,1}=q_{b,2}=q_b$，区域 1 和 2 的积压节点数量是 $n_{b,1}$和 $n_{b,2}$，对区域 1 和 2 中的节点，成功传输的概率为

$$P_{\text{succ},1}=P_{a,1}(1,n_{b,1})P_{a,2}(0,n_{b,2})P_{b,1}(0,n_{b,1})P_{b,2}(0,n_{b,2})+P_{a,1}(0,n_{b,1})P_{a,2}(0,n_{b,2})P_{b,1}(1,n_{b,1})P_{b,2}(0,n_{b,2}) \tag{8-24}$$

$$P_{succ,2} = P_{a1}(0,n_{b,1})P_{a2}(1,n_{b,2})P_{b,1}(0,n_{b,1})P_{b,2}(0,n_{b,2}) + P_{a1}(0,n_{b,1})P_{a2}(0,n_{b,2})P_{b,1}(0,n_{b,1})P_{b,2}(1,n_{b,2}) \quad (8\text{-}25)$$

根据式（8-24）和式（8-25），当q_b很小时，在区域i中节点的成功传输概率可估计为

$$P_{succ,i} \approx [(n_i - n_{b,i})q_{a,i} + n_{b,i}q_b]\mathrm{e}^{-\Lambda(n_{b,1},n_{b,2},n_1,n_2)} \quad (8\text{-}26)$$

令在区域i中增加的积压节点数量是$D_i=(n_i-n_{b,i})\ q_{a,i}-P_{succ,i}$。为让时隙ALOHA稳定，在区域$i$中增加的平均积压节点的数量必须是0，即$\mathrm{E}[(n_i-n_{b,i})q_{a,i}]=\mathrm{E}[P_{succ,i}]$。因此，区域$i$的平均吞吐量为

$$\mathrm{E}[P_{succ,i}] = \mathrm{E}[(n_i - n_{b,i})q_{a,i}] = (n_i - \bar{n}_{b,i})q_{a,i} \quad (8\text{-}27)$$

从Little公式中，我们知道“延时（等待时间）”=“平均系统时间”-“服务时间”。因此，区域i的节点的平均延时为

$$\overline{W}_i = \frac{(n_i - \bar{n}_{b,i})q_{a,i} + \bar{n}_{b,i}}{(n_i - \bar{n}_{b,i})q_{a,i}} - 1 = \frac{\bar{n}_{b,i}}{(n_i - \bar{n}_{b,i})q_{a,i}} \quad (8\text{-}28)$$

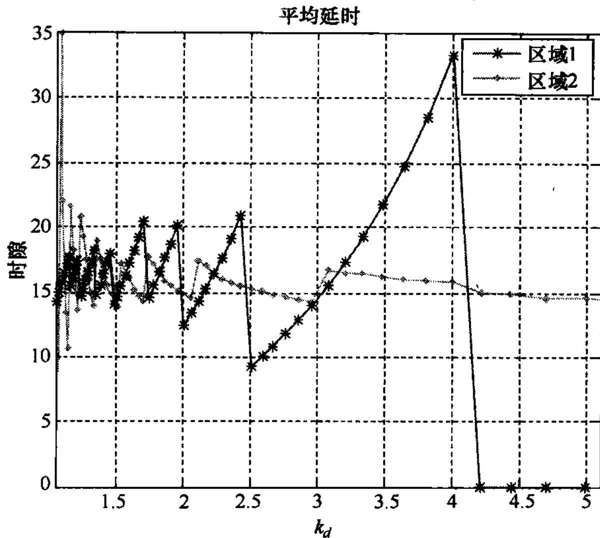


图8-27　区域1和2的平均延时（$q_{b,1}=q_{b,2}=0.05$）

正如预期中的，在区域1和2对于$N=40$和$d_2=80$，平均延时为$q_{a,1}=0.02$和$q_b=0.05$，如图8-27所示。注意，我们仅关注要满足稳定和最大化吞吐量限制条件的情况。在图8-27中，当k_d趋近1时，为了不干扰区域1的节点，数据包到达区域2节点的到达率应该很小。结果是，区域2的节点的吞吐量也小了。在这个情况下，如果区域2中存在积压数据包，它很少有机会再被传输。因此，区域2的节点的平均延时增加了。

相反的，当k_d变大时，区域2中数据包到达率增加了。因此，区域2的节点

的平均延时减少了。图 8-27 中的锯齿现象是因为我们取积压节点数量的期望值为整数。另外，当 k_d增大时，区域 1 的范围减小，因此区域 1 中的节点数量减少。在这个情况下，区域 1 的节点的等待时间将会很大，因为式（8-28）中的分母减小了。一旦 k_d超过 4，区域 1 中无节点，所以等待时间变为 0。同理也可来解释区域 2 中的节点等待时间。

为了完善速率-距离自适应的 ALOHA 系统的分析，我们研究有无限节点的例子。假设在区域 1 和 2 新到达的数据包被认为是一到达就积压了，然后将在下一个时隙被传输。在这个情况下，成功传输的概率为

$$\begin{aligned} P_{\text{succ}} &= n_{b,1}q_{b,1}(1-q_{b,1})^{n_{b,1}-1}(1-q_{b,2})^{n_{b,2}} \\ &\quad + n_{b,2}q_{b,2}(1-q_{b,1})^{n_{b,1}}(1-q_{b,2})^{n_{b,2}-1} \end{aligned} \tag{8-29}$$

令$\partial P_{\text{succ}}/\partial q_{b,1}$和$\partial P_{\text{succ}}/\partial q_{b,2}$为 0，从而最大化$\partial P_{\text{succ}}$。我们能得到 $q_{b,1}=q_{b,2}=1/(n_{b,1}+n_{b,2})$。这意味着积压用户的数量，以及它们在哪个区域并不是重要的。我们所要做的就是取估计系统中的积压用户的总数。总系统到达率是 $\lambda=\lambda_1+\lambda_2$，其中 λ_1是区域 i 的到达率。如果 $\lambda<1/\mathrm{e}$，系统是稳定的。一个伪贝叶斯算法能用来估计积压用户的总数：

$$\hat{n}_{b,k+1}=\begin{cases}\max[\lambda_1+\lambda_2,\hat{n}_{b,k}+\lambda_1+\lambda_2-1], & \text{空闲或成功传输} \\ \hat{n}_{b,k}+\lambda_1+\lambda_2+(\mathrm{e}-2)^{-1}, & \text{碰撞}\end{cases} \tag{8-30}$$

式中，$\hat{n}_{b,k}$是 k 时隙的积压数据包总数的估计。

令 n_i'为区域 i 中传输的数据包的数量。因此，$\mathrm{E}[n_{b,1}+n_1']:\mathrm{E}[n_{b,2}+n_2']=\lambda_1:\lambda_2$。因为所有积压数据包被以同样的概率 $q_b(k)=1/\hat{n}_k$在 k 时隙传输，所以有

$\mathrm{E}[n_1'(k)-n_1'(k-1)\mid n_{b,1}(k)+n_1'(k)-n_1'(k-1)]:$

$\mathrm{E}[n_2'(k)-n_2'(k-1)\mid n_{b,2}(k)+n_2'(k)-n_2'(k-1)]=$

$(n_{b,1}(k)+n_1'(k)-n_1'(k-1)):(n_{b,2}(k)+n_2'(k)-n_2'(k-1))$。取所有值的期望，有

$$\begin{aligned} &\mathrm{E}[n_1'(k)-n_1'(k-1)]:\mathrm{E}[n_2'(k)-n_2'(k-1)] \\ &=\mathrm{E}[n_{b,1}(k)+n_1'(k)-n_1'(k-1)]:\mathrm{E}[n_{b,2}(k)+n_2'(k)-n_2'(k-1)] \end{aligned} \tag{8-31}$$

因为

$$\mathrm{E}[n_1'(k)]:\mathrm{E}[n_2'(k)]=\mathrm{E}[n_1(k)+n_1'(k)]:\mathrm{E}[n_2(k)+n_2'(k)]=\lambda_1:\lambda_2 \tag{8-32}$$

所以有
$$\mathrm{E}[n_{b,1}]:\mathrm{E}[n_{b,2}]=\lambda_1:\lambda_2 \tag{8-33}$$

根据 Little 公式，$\mathrm{E}[n_{b,1}]=\lambda_1\mathrm{E}[W_1]$ 和 $\mathrm{E}[n_{b,2}]=\lambda_2\mathrm{E}[W_2]$，我们得到

$$\mathrm{E}[W_1]=\mathrm{E}[W_2] \tag{8-34}$$

令 $\lambda_2=0.05$ 包/时隙，在区域 1（即用户 1）和区域 2（即用户 2）的积压

的期望值如图 8-28 所示。在图中，与以上公式保持一致，我们发现在区域 1 和 2 的期望积压数据包能根据到达率（例如，$\lambda_1:\lambda_2$）和积压数据包的总数的比例用伪贝叶斯算法估计而获得。区域 1 和 2 的平均延时如图 8-28 所示。我们能发现区域 1 和 2 的平均延时几乎是一样的，而且能从伪贝叶斯算法的平均时间来获得，这显示了比不用速率-距离特性的 ALOHA 更好的吞吐量延时性能。

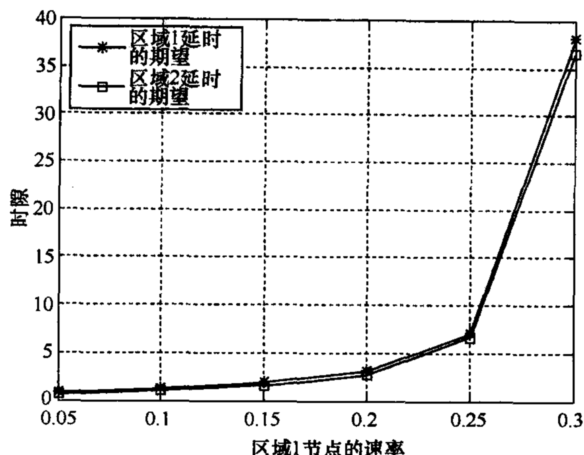


图 8-28 平均延时性能

如果总系统到达率 λ 是已知的，当所有的节点是在整个系统区域均匀的固定分布时，区域 1 和区域 2 的节点到达率分别为 $\lambda_1=\lambda d_1^2/d_2^2$ 和 $\lambda_2=\lambda(1-d_1^2/d_2^2)$。因为系统吞吐量是 $G(n)\mathrm{e}^{-G(n)}$ 和 $G(n)=n_{b,1}q_r+n_{b,2}q_r$，其中 q_r 是重传概率，吞吐量被区域 1 和 2 中积压数据包的数量和积压重传概率影响。如果 $q_{b,1}=q_{b,2}=q_r=1/(n_{b,1}+n_{b,2})$，吞吐量可被最大化。否则，如果 $q_{b,1}=q_{b,2}\neq 1/(n_{b,1}+n_{b,2})$，惟一的影响吞吐量的参数将是重传概率 q_r。

8.4 媒体接入控制的 CSMA

在多认知无线电中支持 DSA 的媒体接入控制协议显然是一个认知无线电网络的关键问题。这里，我们考虑由多个认知通信参数的认知无线电与合法主系统共存而形成的认知无线电网络。在认知无线电上处理 ALOHA 后，它可直接研究 CSMA 协议。为了这个主要目标，各种认知无线电网络的多址接入协议被提出了。这其中，CSMA 是一个简单而有效的多址接入方案，为认知无线电网络提供了一个高吞吐量，而不用去考虑如何进行空间复用。根据在第 5 章中关于无线通信的速率-距离的特性，认知无线电通信也许能够在主系统的传输期间，对于空间复用并在一个可接受的干扰下，同时处理。因此，一个有效的认知无线电网络

的多址接入协议可以融合空间复用能力，而充分地在时域和空域上使用无线电资源。

并发传输的 CSMA 已经在 Ad hoc 和多跳网络中主要作为同构网络，引起很多关注。认知无线电网络中的一个问题是合法主系统的修改操作常常是不可接受的。同时 Wang 等在文献［30］中提出，如果在认知无线电网络中每个基站的精确位置信息已知，在主系统传输期间的认知无线电并发传输反而提高了总吞吐量。通过分析上行链路 SINR 和下行链路 SINR，然后告知 CSMA/CA 精确的位置信息，这是可以达成的。接下来，我们基于速率-距离特性扩展研究范围，提出空间复用传输的 CSMA（称为 CSMA-ST）作为认知无线电网络的多址接入协议。不用要求精确的位置信息，这个常常在认知无线电网络中即使用全球定位系统也难达到，认知无线电仅仅需要信号干扰信噪比去适应已知确定的操作的合法主系统。提出的 CSMA-ST 因此一般允许在 PS-Tx 无位置信息的传输期间进行可能的认知无线电的同时并发包传输。并发传输也许意味着更高的干扰，这引发了认知无线电网络的另外一个关键任务：恰当地控制到主系统的干扰。多址接入协议自身是不足以达到最优协议容量，也不能抑制到主系统的干扰。有功率控制和 MAC 的跨层设计因此是需要用来完善无位置信息要求的任务。

实际操作中，主系统的功能仅是合法设备，其中并不需要将主系统的节点像认知无线电一样还要运行最新的媒体接入控制协议。认知无线电网络中的媒体接入控制应该能够方便合法设备的共存。同时，认知无线电网络媒体接入控制的最终目标是让每个认知无线电在干扰的限制下主系统中断概率低于预定值，能够同时以最大速率传输数据包，还能够最大化认知无线电网络和主系统的吞吐量、认知无线电网络的总吞吐量。回顾传输、距离、每个节点容忍干扰和到其他节点干扰的互相作用，基于功率-速率控制方案和跨层功率-速率控制步骤的信道转换应该便于 CSMA-ST。

无线通信的速率-距离的特性已在 5.4.3 节中讲述了。我们考虑认知无线电网络和主系统的共存情况，如图 8-29 所示，这很像图 8-24。在主系统中的基站和移动站的调制方案和数据速率是基于它们收到的信号干扰信噪比（SINR）。为不失一般性，我们考虑两种速率：高速率和低速率。在主系统中靠近基站的移动站选择一个高速率，反之远离基站的移动站选择一个低速率。在主系统中靠近基站的覆盖边界时，认知无线电的通信链路能够建立起来。一个高速率调制方案也许在两个认知无线电中是可以阻止主系统中来自传输站的干扰。自传输站到接收站的干扰也许是对于主系统中被接收站采用的调制方案是可接受的。

8.4.1 载波监听空间复用传输的多址接入

为了减轻隐藏终端问题和提高吞吐量，提出的 CSMA-ST 采用四路握手程

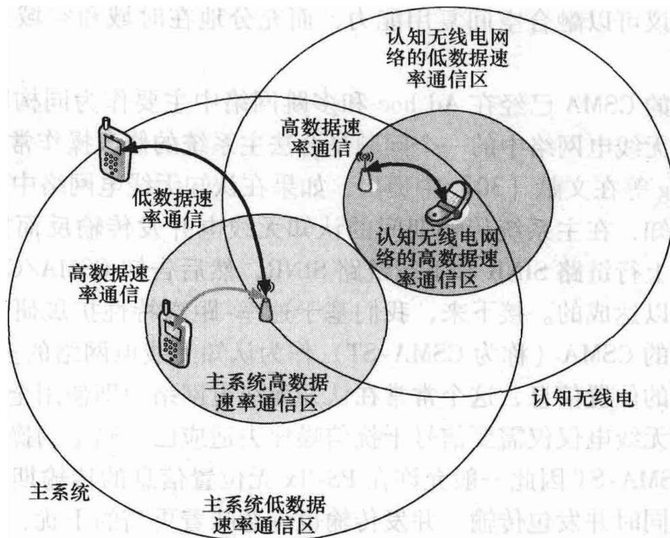


图 8-29 认知无线电网络中的 CSMA-ST 的网络情景
（这里的认知无线电网络由认知无线电组成）

序[18]。认知无线电网络包括了一个 BS 或 Ad hoc 网络和多认知无线电的报文接收器去尝试传输数据包到基站。合法主系统采用传统的 CSMA 类协议（即碰撞避免的 CSMA，CSMA/CA）。不同与碰撞的 CSMA，它考虑了通道要么被占据（PS 的一个传输站在传输数据包）或者空闲状态（PS 的所有传输站不传输数据包），提出的 CSMA-ST 考虑通道为：1）占据状态（PS 的一个传输站在传输数据包对认知无线电无传输机会）；2）忙状态（PS 的一个传输站在传输数据包，但是认知无线电能找机会传输）；3）空闲状态。

一个传输机会意味着认知无线电能为传输识别一个可行的功率和速率，这需要满足：

■ 从 CR 传输到 PS 的 RS 的干扰在一个可接受的水平上可控。

■ 从 PS 的 TS 的干扰是可以解决的，即 CR 的误码率（BER）是维持不变的。

当通道被 CR 和 PS 的 TS 感知到空闲时，CR 等待一个周期 τ_s，因此 PS 的 TS 能接入信道优先于 CR 的信道接入。两个统计参数可用来描述 PS 的 TS 行为特征（见图 8-30a）：

■ 在 CR 的等待期间结束前 PS 的所有 TS 的总成功传输概率是 q_p；

■ 在 CR 的等待期间结束后，剩余的数据传输时间是 τ_x。

在等待期间结束，当 CR 尝试传输数据包时信道被认为是忙状态，RTS 消息被送出到认知无线电网络的 BS，同时 PS 的 TS 在传输竞争信道。如果信道竞争是成功的，带有可行的功率和数据率处理的并发数据包传输如图 8-30a 所示。否

则，CR 等待直到信道空闲，如图 8-30b 所示，其中这些数据包认为是积压的。

图 8-30　认知无线电网络的 CSMA-ST 流程（数据传输结束时 ACK 消息被包含了）

存在着有一个并发传输的机会，其中对于 CR 的每次成功的信道竞争，只有一个数据包能够被传输。为了更高效的使用无线电资源，在 PS 的多 TS 中，CSMA-ST把一个长期积压时间或者连续的数据包碰撞看作一个“频谱空洞”。如果在 CR 的等待期间结束前，PS 的成功包传输不发生，它们能竞争信道和传输

数据包。在这个情况下，PS 的 TS 认为信道是被占据的，不再传输数据包，直到下一个信道空闲时刻到来，如图 8-30c 所示。当信道空闲、在 CR 的等待期间结束时 CR 的 RTS 消息能被传输（这也是一个频谱空洞），如图 8-30d。

从 CR 发射机到 PS 的一种不可接受的干扰可建模成 PS 的中断概率，即从 CR 发射机到 PS 的 RS 的干扰概率大于 I_{acc}。中断概率将不超过预定的 P_{outage}，表示为

$$Pr\{I_{ps} > I_{acc}\} \leqslant P_{outage} \tag{8-35}$$

其中，I_{ps}是从 CR 发射机到 PS 接收机的干扰功率。这是 CR 传输的干扰限制。

基于以上的特点，所提出的 CSMA-ST 协议的详细操作阐述如下：

1. 对于每个 CR 发射机，当 CR 发射机在等待期间尝试去传输数据包时，在等待期间结束时它感知信道并转到第二步。如果不是等待期间，CR 发射机等待直到信道变为空闲。

2. 当信道变为空闲，CR 传输机等待一段时间，τ_s，然后在等待时间结束时感知信道。

a）如果在等待时间结束时从 PS 的传输站到 CR 发射机的干扰功率小于一个预定的阈值，CR 发射机送出 RTS 消息到认知无线电网络的 BS 去竞争信道。

b）否则，CR 发射机将等待直到信道空闲，然后重复第一步和第二步以一个合适的重传概率再发送 RTS。

3. 如果一个 RTS 被 CR 接收机正确地接收，基于接收到的 RTS 的信号干扰信噪比（SINR），CR 接收机计算可行的传输功率和速率。如果一个可行的传输功率和速率是可获得的，这个信息通过回复 CTS 被传回 CR 发射机。否则，CTS 不被回复。

4. 如果相应的 CTS 被 CR 发射机接收到，根据 CTS 的嵌入信息它发射带有功率和速率的数据包。否则，CR 发射机将等待直到信道空闲，然后重复第一步和第二步以一个合适的重传概率再发送 RTS。

5. 当数据包被 CR 接收机正确地接收到，就会回复一个 ACK，因为动态数据链路可用于 CR。

8.4.2 CSMA-ST 分析

为了分析 CSMA-TS，我们首先假设：

1. 不失一般性，CR 可以有两种速率，高速率和低速率。接近和远离认知无线电网络的 BS 分别有高和低数据速率。

2. 令数据包到达 N_h个高速率 CR 和 N_l个低速率 CR 有独立的平均到达率，分别为 λ_h和 λ_l泊松过程。

3. 如果一个数据包在等待期间到达一个非积压的 CR（即无积压数据包的 CR），一个 RTS 消息能被传输到认知无线电网络的 BS 去竞争通道。令 $q_{ah} = (1 - e^{-\lambda_h \tau_s})\ q_{Rh}$和 $q_{al} = (1 - e^{-\lambda_l \tau_s})\ q_{Rl}$分别代表 RTS 每个非积压的 CR 靠近和远离认知

无线电网络的 BS 的传输概率。q_{Rh}和 q_{Rl}分别是从 PS 的传输站对于高速率和低速率 CR 发射机的干扰功率不超过预设的阈值的概率。

4. 如果一个数据包到达一个非积压 CR，但是信道是非空闲或者 CR 正在传输数据包，这个数据包被认为是一个积压数据包。

5. 每个 CR 假设没有缓冲器，因此新数据包到达积压 CR（有积压数据包的 CR）就被丢掉了。

6. 每个积压 CR，在等待时间结束时如果 RTS 消息被允许传输，以概率 q_{Rh}和 q_{Rl}分别为高速率和低速率 CR 再发送 RTS 消息。

7. 数据包为高速率和低速率 CR 的传输时间分别是 T_h和 T_l，且有 $T_h \leqslant T_l$。

8. T_h和 T_l都比剩余的 PS 的 TS 数据传输周期 τ_x要短。

一个二维有限状态马尔可夫链，如图 8-31 所示，用于性能分析。状态索引（n_{bh}，n_{bl}）代表着积压的高速率和低速率 CR 的数量。当既没有新数据包或没有 CR 的积压数据包尝试被传输，PS 的所有 TS 也没有成功传输时，在等待期间结束时状态转移了。否则，在忙（占有）信道期间结束时状态转移。

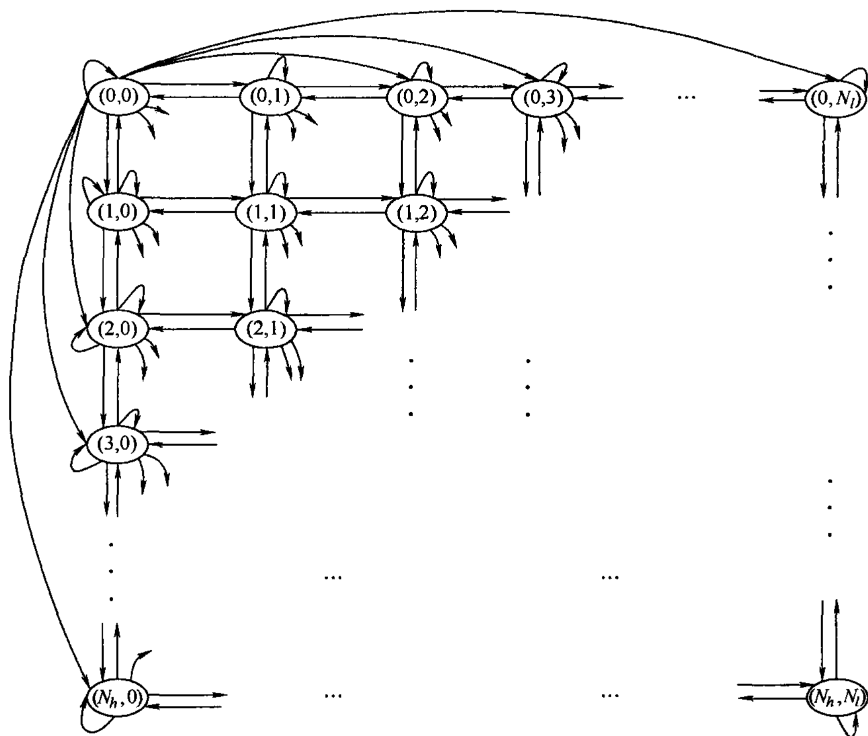


图 8-31 二维有限状态马尔可夫链，用（n_{bh}，n_{bl}）作为用来分析积压的高速率和低速率 CR 的数量的状态索引

令τ_{RTS}和τ_{CTS}分别为RTS和CTS消息的接收时间包括最差的数据包传输延时。令$g_h(n_{bh})=(N_h-n_{bh})q_{ah}+n_{bh}q_{bh}$和$g_l(n_{bl})=(N_l-n_{bl})q_{al}+n_{bl}q_{bl}$分别为在等待期间结束时总RTS高速率和低速率CR的传输概率。如果所有PS的TS在等待期间结束前无成功传输的概率是（$1-q_p$），有4个状态在二维有限状态马尔可夫链（FSMC）中：

1. τ_s（既没有新数据包或没有重传的积压数据包尝传输的情况）概率为$e^{-(g_h(n_{bh})+g_l(n_{bl}))}$

2. $\tau_s+\overline{T}_h$（高速率CR成功传输的情况，$\overline{T}_h=T_h+\tau_{RTS}+\tau_{CTS}$）概率为$g_h(n_{bh})e^{-(g_h(n_{bh})+g_l(n_{bl}))}$

3. $\tau_s+\overline{T}_l$（低速率CR成功传输的情况，$\overline{T}_l=T_l+\tau_{RTS}+\tau_{CTS}$）概率为$g_l(n_{bl})e^{-(g_h(n_{bh})+g_l(n_{bl}))}$

4. $\tau_s+\beta$（CR碰撞的情况，$\beta=\tau_{RTS}+\tau_{CTS}$）概率为$1-e^{-(g_h(n_{bh})+g_l(n_{bl}))}-(g_h(n_{bh})e^{-(g_h(n_{bh})+g_l(n_{bl}))})$，或者因为PS的TS以概率$q_R$的干扰，RTS接收失败的情况。

在CR的等待期间结束前PS的TS以概率q_p有一次成功的传输，状态转移周期是X_p，$X_p=\tau_s+\tau_x$。因此，状态转移的期望周期是

$$\begin{aligned}\Gamma_c(n_{bh},n_{bl})=(1-q_p)[\tau_s(1+q_R)+(\overline{T}_{hg_h}(n_{bh})+\overline{T}_{lgl}(n_{bl}))\cdot\\ e^{-(gh(n_{bh})+gl(n_{bl}))}+(q_R+1-e^{-(gh(n_{bh})+gl(n_{bl}))})\\ -(gh(n_{bh})+gl(n_{bl}))e^{-(gh(n_{bh})+gl(n_{bl}))})\beta]+q_pX_p\end{aligned} \quad (8\text{-}36)$$

8.4.2.1 CSMA-ST的吞吐量

高速率和低速率CR的每次状态转移成功传输的期望值分别是$g_h(n_{bh})e^{-(g_h(n_{bh})+g_l(n_{bl}))}$和$g_l(n_{bl})\ e^{-(g_h(n_{bh})+g_l(n_{bl}))}$。通过获得二维FSMC的$n_{bh}$和$n_{bl}$的稳态分布，使用提出的CSMA-ST认知无线电网络的吞吐量表示为

$$S_c(n_{bh},n_{bl})=\frac{(gh(n_{bh})+gl(n_{bl}))e^{-(gh(n_{bh})+gl(n_{bl}))}}{\Gamma_c(n_{bh},n_{bl})} \quad (8\text{-}37)$$

这是离开每个状态转移的数据包的数量。

8.4.2.2 CSMA-ST的平均数据包延时

平均数据包延时分析可以从修改一个典型的CSMA的平均延时获得。高速率和低速率CR的平均数据包延时可估计为

$$\begin{aligned}W_{\text{high-rate}}\approx\frac{1}{2(1-(\overline{\lambda}_h+\overline{\lambda}_l)E[t])}\left[(1-q_p)\left(\frac{\overline{\lambda}_h(\tau_s+\overline{T}_h)}{\overline{\lambda}_h+\overline{\lambda}_l}+\frac{\overline{\lambda}_l(\tau_s+\overline{T}_l)}{\overline{\lambda}_h+\overline{\lambda}_l}\right)+q_pX_p\right]^2\\ +2[E[t]-(\tau_s+\overline{T}_h)]\\ W_{\text{low-rate}}\approx\frac{1}{2(1-(\overline{\lambda}_h+\overline{\lambda}_l)E[t])}\left[(1-q_p)\left(\frac{\overline{\lambda}_h(\tau_s+\overline{T}_h)}{\overline{\lambda}_h+\overline{\lambda}_l}+\frac{\overline{\lambda}_l(\tau_s+\overline{T}_l)}{\overline{\lambda}_h+\overline{\lambda}_l}\right)+q_pX_p\right]^2\\ +2[E[t]-(\tau_s+\overline{T}_l)]\end{aligned} \quad (8\text{-}38)$$

式中，E[t] 是吞吐量的倒数，可解释为对认知无线电网络的离开数据包的期望要求时间。$\overline{\lambda}_h=(N_h-n_{bh})\ \lambda_h$和$\overline{\lambda}_l=(N_l-n_{bl})\ \lambda_l$分别是高速率和低速率 CR 的累计到达率。

8.4.2.3　不同速率下的 CSMA-ST 的性能

图 8-32 所示为在低速率认知无线电的不同数据包传输时间 T_l（T_h 的倍数）下 CSMA-ST 的认知无线电网络的吞吐量和平均数据包延时。

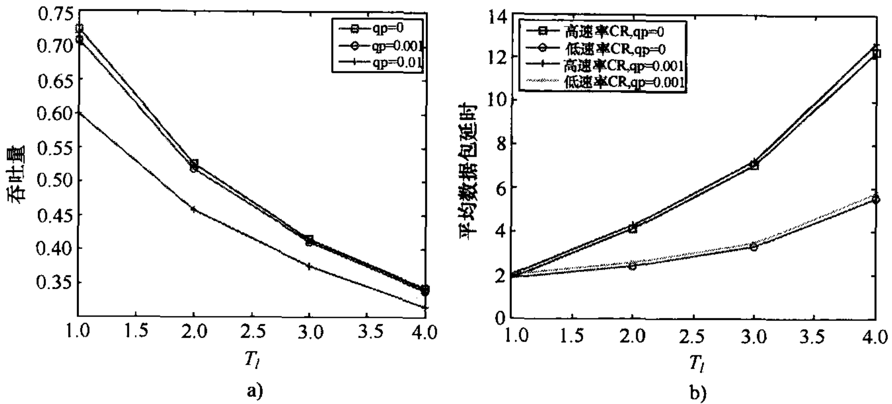


图 8-32　CSMA-ST 的吞吐量和延时

a）不同 T_c 情况下的 CSMA-ST 吞吐量（$T_h=1$，$\tau_s=0.12$，$\tau_R TS=0.02$，$\lambda_h=\lambda_1=0.1$，$N_h=N_1=10$，$q_{bn}=q_{b1}=0.04$，$q_R=0.01$，$q_{RH}=q_{Ri}=0.99$，$X_p=10.$）　b）不同 T_c 情况下的 CSMA-ST 平均数据包延时（$T_h=1$，$\tau_s=0.12$，$\tau_R TS=0.02$，$\lambda_h=\lambda_1=0.1$，$N_h=N_1=10$，$q_{bn}=q_{b1}=0.04$，$q_R=0.01$，$q_{RH}=q_{Ri}=0.99$，$X_p=10.$）

我们指出减少低速率认知无线电的包传输时间（即增加速率）能增加吞吐量和减少认知无线电网络的平均包延时。然而，高速率传输依靠高信号干扰信噪比，也依靠高传输功率。数据速率不能无限制地增加，因为认知无线电阻止到 PS 的不可接受的干扰。为了在干扰限制下最大化吞吐量和最小化认知无线电网络的平均包延时，认知无线电应该使用最大的传输数据速率去并发传输数据包。这个通信策略也可用于在干扰限制下最大化空间复用效率。

8.4.2.4　总吞吐量分析

细心的读者也许想要知道认知无线电网络的吞吐量的增加是否意味着总吞吐量的增加。为了验证，我们分析一下总吞吐量。

考虑如图 8-33 所示的共存网络拓扑，我们把靠近和远离认知无线电接收机的认知无线电发射机记为 CR_{near}和 CR_{far}。距离矢量 [a，b，c，d，e，f] 用于代表 PS 的认知无线电和主系统站的距离。在这个分析中，考虑功率谱密度为 $N_0/2$ 的加性高斯白噪声具有路径损失的一个通道。CR_{near}和 CR_{far}的数据速率（归一化

到信道带宽）和 PS 的 TS 表示为

$$c\mathrm{CR}_{\mathrm{near}}=\log_2\left(1+\omega\frac{P_{\mathrm{CR}_{\mathrm{near}}}a^{-\alpha}}{P_{\mathrm{PS}}f^{-\alpha}+N_0B}\right) \tag{8-39}$$

$$c\mathrm{CR}_{\mathrm{far}}=\log_2\left(1+\omega\frac{P_{\mathrm{CR}_{\mathrm{far}}}b^{-\alpha}}{P_{\mathrm{PS}}f^{-\alpha}+N_0B}\right) \tag{8-40}$$

$$c_{\mathrm{PS}}=\log_2\left(1+\omega\frac{P_{\mathrm{PS}}e^{-\alpha}}{(\eta P_{\mathrm{CR}_{\mathrm{near}}}c^{-\alpha}+(1-\eta)P_{\mathrm{CR}_{\mathrm{far}}}d^{-\alpha})+N_0B}\right) \tag{8-41}$$

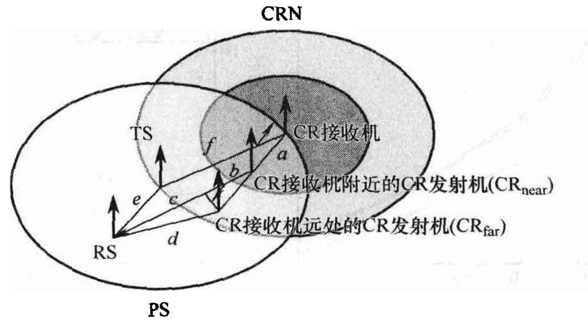


图 8-33 一个典型的认知无线网络和主系统共存的网络拓扑

式中，P_{PS}是 PS 的 TS 的传输功率，ω 是关于 BER 的常量。α 是路径损失指数。$P_{\mathrm{CR}_{\mathrm{near}}}$和 $P_{\mathrm{CR}_{\mathrm{far}}}$分别是 $\mathrm{CR}_{\mathrm{near}}$和 $\mathrm{CR}_{\mathrm{far}}$的传输功率。$B$ 是信道带宽。$\mathrm{CR}_{\mathrm{near}}$和 $\mathrm{CR}_{\mathrm{far}}$的成功信道竞争概率是 η 和（$1-\eta$），其中 $0\leqslant\eta\leqslant1$。

我们估计在等待期间结束时高速率和低速率的认知无线电的累计 RTS 消息传输概率为 g_h和 g_l。认知无线电网络的吞吐量 S_{CRN}估计为

$$S_{\mathrm{CRN}}\approx\frac{(g_h+g_l)\mathrm{e}^{-(g_h+g_l)}}{\Gamma_c} \tag{8-42}$$

式中，$\Gamma_c\approx(1-q_p)[\tau_s(1+q_R)+\overline{T}_hg_h\mathrm{e}^{-(g_h+g_l)}+\overline{T}_l\,g_l\mathrm{e}^{-(g_h+g_l)}$
$+(1-\mathrm{e}^{-(g_h+g_l)}-(g_h+g_l)\mathrm{e}^{-(g_h+g_l)}+q_R)\beta]+q_pX_p$

PS 的吞吐量（归一化到认知无线电传输机无干扰的、仅有高斯白噪声（AWGN）的数据速率）估计为

$$S_{\mathrm{PS}}\approx\frac{q_p\tau_x(c_{\mathrm{PS}}/c_{\mathrm{PS_AWGN}})}{\Gamma_c}$$
$$c_{\mathrm{PS_AWGN}}=\log_2(1+\omega P_{\mathrm{PS}}\mathrm{e}^{-\alpha}/(N_0B)) \tag{8-43}$$

S_{PS}可认为是认知无线电存在时的吞吐量。因此，总吞吐量为

$$S_{\mathrm{PS+CRN}}\approx\frac{(g_h+g_l)\ \mathrm{e}^{-(g_h+g_l)}+q_p\tau_x\ (c_{\mathrm{PS}}/c_{\mathrm{PS_AWGN}})}{\Gamma_c} \tag{8-44}$$

8.4.2.5　俘获效应和并发传输条件

如果超过一个并发传输到达接收机，强信号被接收机接收，而弱信号被认为是噪声或干扰。这就是无线电通信的俘获效应。我们采用一个简单但是广泛采用的俘获效应模型。在这个模型中，俘获效应发生，当

$$P_{\mathrm{CR_{near}}}a^{-\alpha} \geqslant (P_{\mathrm{PS}}f^{-\alpha} + N_0B),$$
$$P_{\mathrm{CR_{near}}}b^{-\alpha} \geqslant (P_{\mathrm{PS}}f^{-\alpha} + N_0B), \tag{8-45}$$
$$P_{\mathrm{PS}}e^{-\alpha} \geqslant [(\eta P_{\mathrm{CR_{near}}}c^{-\alpha} + (1-\eta)P_{\mathrm{CR_{far}}}d^{-\alpha}) + N_0B]$$

这是并发传输的条件（即 CR 接收机的 SINR 和 PS 的接收站超过 0dB）

8.4.2.6　总吞吐量的性能

在图 8-34 中，我们数值上画出了 CRN、PS 和总吞吐量在三种情况下的吞吐量。当并发传输的条件满足时，我们分析了总吞吐量，距离矢量选为 D_1，如图 8-34 所示。假设 P_{PS}和噪声功率固定，我们逐渐的减少 $P_{\mathrm{CR_{near}}}$和 $P_{\mathrm{CR_{far}}}$，因此 PS 接收站的 SINR 逐渐增加，如图 8-34 所示。我们观察到，在给定的流量下，认知无线电发射机填满了频谱空洞，但是无并发传输导致了认知无线电网络的吞吐量只有 0.22。在这个情况下，PS 的吞吐量大约是 0.1，总吞吐量大约是 0.32。认知无线电的并发数据包传输时最大的总吞吐量为 0.55。

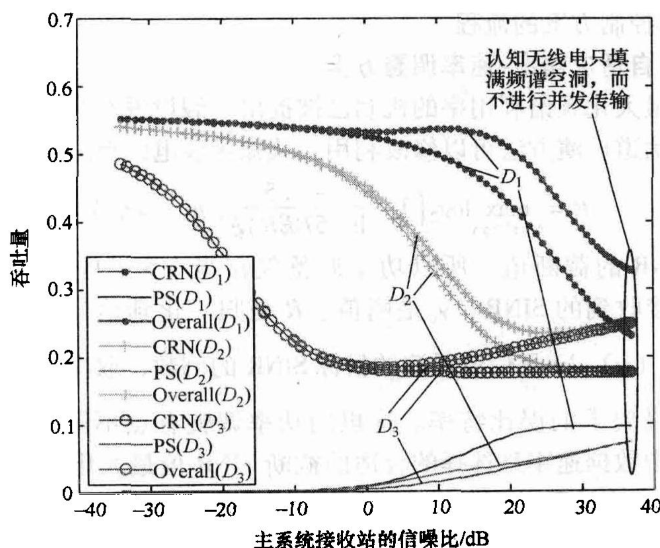


图 8-34　认知无线网络、主系统和总吞吐量的吞吐量（$g_h = g_l = 0.4$，$\tau_s = 0.12$，$x_p = \tau_x + \tau_s = 10$，$T_h = 1$，$T_l = T_h c_{\mathrm{CR_{near}}}/c_{\mathrm{CR_{far}}}$，$\eta = 1/2$，$B = 200\mathrm{kHz}$，$w = 0.28$，$q_R = 0.01$，$q_{Rh} = q_{Rl} = 0.99$，$d = 2$，$D_1 = [20, 40, 300, 240, 50, 260]$，$D_2 = [20, 50, 120, 100, 50, 120]$，$D_3 = [100, 120, 50, 20, 80, 60]$）

我们下一步缩短认知无线电网络和PS距离（因此认知无线电网络和PS的干扰增加了），但是同时传输的条件还保持。距离矢量选为D_2，如图8-34所示。CR发射机填满了频谱空洞，但是非并发传输导致了0.26的总吞吐量。然而，并发数据包的传输导致了最大的吞吐量只有0.47的总吞吐量。当并发传输条件改变时，距离矢量选择为D_3，如图8-34所示。在这个情况下，并发传输引进了在PS和认知无线电网络间的不可接受的干扰。因此，总吞吐量减少了。

总之，我们观察到：1）认知无线电网络吞吐量的增加提高了总吞吐量。2）PS的吞吐量的牺牲可以得到更多的总吞吐量，如果并发传输条件是满足的。实际上，在（i）中认知无线电网络不保证严格的限制条件。我们还能观察到最大化干扰限制下认知无线电网络的吞吐量也最大化了同样限制条件的总吞吐量。为了这个目的，我们在下面章节中提出一个跨层的功率-速率控制方案。

8.4.3 跨层功率-速率控制方案

为了开发一个完整的设计在干扰的限制下优化MAC，我们首先引入了功率-速率调整方案去最大化数据速率（也是频谱效率）。然后，在认知无线电的干扰限制下最大化的传输功率推导为功率调整的SINR的切断阈值。最后，我们总结跨层功率-速率控制方案的流程。

8.4.3.1 自适应功率-速率调整方案

很多关于最大化频谱利用率的机制已被提出。假设没有自适应M-QAM星座图大小限制，信道反演方法可以修改利用，认知无线电的自适应功率-速率为

$$R=\max_{\gamma',\gamma'>\gamma_0}\log_2\left(1+\frac{-1.5}{\ln(5BER)\xi}\right)p(\gamma>\gamma') \tag{8-46}$$

式中，γ'是SINR的截断值，所以功率调整仅在当$\gamma>\gamma'$且概率是p（$\gamma>\gamma'$）时进行，γ是接收到的SINR。γ_0是阈值。R是归一化到信号带宽的数据速率，$\xi\triangleq\int_{\gamma'}^{\infty}(1/\gamma)p(\gamma)\,\mathrm{d}\gamma$是信道反演的目标SINR的倒数，取决于认知无线电的平均功率，BER是要求的误比特率。这里的功率调整有点不同于参考文献[18，19]所述，其中数据速率是选择的合适的截断SINR值最大化。因为认知无线电将阻止到PS的不可接受的干扰，在式（8-46）的数据速率仅在截断SINR值超过阈值γ_0时才被最大化。我们在下面章节中推导γ_0。

8.4.3.2 截断SINR阈值

为了获得γ_0，我们应该首先获得在干扰限制下认知无线电的最大传输功率。这个传输功率是最大的可允许传输功率Ψ，表示为

$$\Psi=I_{\mathrm{acc}}\left[10\alpha\log_{10}(d_{P_{\mathrm{outage}}}/d_0)+q(\gamma)_{P_{\mathrm{outage}}}\right] \tag{8-47}$$

式中，$d_{P_{outage}}$是距离，因此$\int_0^{d_{P_{outage}}} d(r)\,dr \leq P_{outage} q(\gamma)_{P_{outage}}$是一个值，$\int_0^{q(\gamma)P_{outage}} q(\gamma)\,d\gamma \leq P_{outage} q(\gamma)$。$q(\gamma)$ 是接收到的 SINR 的分布。r 和 $d(r)$ 是欧式距离，分布于认知无线电的任一个发射机和 PS 的接收站之间[30]。d_0 是参考距离。那么，获得截断 SINR 阈值为

$$\gamma_0 = \Psi_\rho / (N_0 B + I) \tag{8-48}$$

式中，ρ 是信道增益。I 是 PS 发射站的干扰功率。为了最大化数据速率，认知无线电应该搜索一个可行的超过 γ_0的截断 SINR 去最大化式（8-48）。

8.4.3.3　**跨层功率-速率控制流程**

跨层功率-速率控制方案的流程总结如下：

1. CR 使用根据式（8-47）最大的可允许的传输功率发送 RTS 到 CR 接收机去竞争信道。

2. 当 RTS 消息被 CR 接收机接收到了，CR 接收机根据式（8-48）决定 γ_0。

3. 如果在 RTS 的 SINR 不超过 γ_0，CTS 消息不回复。否则，CR 接收机搜索一个可行的超过 γ_0的截断 SINR 去最大化式（8-46）。

a）如果 RTS 的 SINR 超过截断 SINR，反转信道到目标 SINR，$1/\xi$ 和式(8-46)的数据速率的传输功率被携带在回复给 CR 发射机中传输。

b）否则，CTS 消息不给回复。

因为认知无线电网络的隐藏终端问题，认知无线电感知活跃的 PS 接收站比感知活跃的 PS 发射站更加困难。因此，使用最大允许传输功率的 CR 发射机发送 RTS 是很必要的，即使不是为了并发传输。

8.4.4　性能评估

为了进一步考虑性能上不可靠的无线信道的影响，我们建立起一个有或没有自适应功率-速率调整的多址接入方案之间的 BER 和数据速率的关系。然后，四种多址接入方案的理论吞吐量描述成允许适当的性能评估和比较。

我们从有或没有自适应功率-速率调整的多址接入方案之间的数据速率和 BER 的关系的推导开始，

$$\frac{BER_2}{BER_1} \approx (5BER_1)^{\frac{2^{c_2}-1}{2^{c_1}-1}} - 1 \tag{8-49}$$

式中，BER_1 是具有功率和速率调整的数据速率 C_2 的多址接入方案的 BER。这个 BER 被认为是不影响吞吐量。BER_2 是无功率和速率调整的数据速率 C_1 的多址接入方案的 BER。

假设靠近认知无线电网络的基站的认知无线电有可接受的 BER，但是远离认知无线电网络的基站的认知无线电有传输错误，根据式（8-49），如果功率和

速率控制调整是不可用的，我们简化地使用 BER 去估计数据包错误率。多址接入方案 2、3 和 4 是时隙的。数据包仅在每个时隙的开头传输。除了特别声明外，在这节中的定义也和前几节中的定义一样。我们现在描述多址接入方案的流程。

方案 1：无功率调整的 CSMA（CSMA 非-PRA）：认知无线电采用 CSMA-ST，但是仅通过频谱空洞传输数据包。物理层不调整传输功率和数据率，因此 BER 不能被保持在一个要求的值。吞吐量为

$$S_{\mathrm{CSMA_nonST}} \approx \frac{(1-q_p)\left(g_h + g_l \cdot (5BER_1)^{-\left(\frac{2^{c_2}-1}{2^{c_1}-1}-1\right)} \mathrm{e}^{-(g_h+g_l)}\right)}{\Gamma_c} \tag{8-50}$$

式中，远离认知无线电网络的基站的认知无线电的成功传输概率为

$$\left(g_l \cdot (5BER_1)^{-\left(\frac{2^{c_2}-1}{2^{c_1}-1}-1\right)}\right)\mathrm{e}^{-(g_h+g_l)}$$

认知无线电仅有一个数据速率：$\overline{T}_h = \overline{T}_l$，在 Γ_c 中。

方案 2：具有功率调整（RA-PRA）的随机接入：认知无线电动态地以时隙 ALOHA 的方式接入信道。即在一个时隙开始且信道空闲时，靠近或远离认知无线电网络中的基站的认知无线电分别以概率 g_h 和 g_l 传输数据包。功率-速率调整目标是保持无线信道要求的 BER，但不是为了并发传输。我们估计吞吐量为

$$S_{\mathrm{CR_AC}} \approx \frac{(g_h+g_l)\mathrm{e}^{-(g_h+g_l)}}{\left\{\begin{array}{l}(1-\mathrm{e}^{-g_l})T_l + (1-\mathrm{e}^{-g_h})\mathrm{e}^{-g_l}T_h \\ +q_f T_S \mathrm{e}^{-(g_h+g_l)} + \mathrm{e}^{-(g_h+g_l)}(1-q_f)X_p\end{array}\right\}} \tag{8-51}$$

式中，q_f 是 PS 的所有 TS 在一个 T_S 的时隙都未成功传输的概率。

方案 3：无功率调整的随机接入（RA 非-PRA）：当 PS 的所有 TS 不传输数据包时，认知无线电以 s-ALOHA 的方式接入信道。物理层不调整传输功率和数据速率。这个方案的吞吐量为

$$S_{\mathrm{CRnon_AC}} \approx \frac{\left(g_h + g_l \cdot (5BER_1)^{-\left(\frac{2^{c_2}-1}{2^{c_1}-1}-1\right)}\right)\mathrm{e}^{-(g_h+g_l)}}{\{T_{\mathrm{CR}}(1-\mathrm{e}^{-(g_h+g_l)}) + q_f T_s \mathrm{e}^{-(g_h+g_l)} + \mathrm{e}^{-(g_h+g_l)}(1-q_f)X_p\}} \tag{8-52}$$

式中，T_{CR} 是认知无线电的数据包传输时间。

方案 4：无认知的随机接入（RA 无认知）：认知无线电以 s-ALOHA 的方式随机接入信道，不管 PS 的 TS 是否传输时间包。物理层不调整传输功率和数据速率。在这个情况下，认知无线电的并发传输数据包和 PS 的 RS 互相干扰（数据包碰撞）。这个方案的吞吐量为

$$S_{\mathrm{MACnon_cognition}} \approx \frac{\left(g_h + g_l \cdot (5BER_1)^{-\left(\frac{2^{c_2}-1}{2^{c_1}-1}-1\right)}\right)\mathrm{e}^{-(g_h+g_l)}}{\{T_{\mathrm{CR}}(1-\mathrm{e}^{-(g_h+g_l)}) + q_f T_s \mathrm{e}^{-(g_h+g_l)} + \mathrm{e}^{-(g_h+g_l)}(1-q_f)X_p\}} \tag{8-53}$$

值得注意的是 RA 非认知有像 RA 非-AC 一样的成功数据包离开速率。那

么，它们有同样的理论吞吐量。

我们使用了认知无线电网络的远离基站的认知无线电的数据包传输时间 T_l 去表示 PS 的干扰水平和信道质量（接收到的 SINR）。设定参数如下：$q_p=0.01$，$X_p=10$，$\tau_{RTS}=\tau_{CTS}=0.02$，$\tau_S=0.12$，$T_{CR}=T_S=1$，$T_h=1$，$BER_1=10^{-3}$，$q_R=0.01$，$q_{RH}=q_{RI}=0.99$，$q_f=0.99$。图 8-35 所示为在认知无线电网络的基站的高 SINR 条件下，那么远离认知无线电网络（$T_l=2$）的基站的认知无线电会有高数据速率。所提出的方案有一个 49% 到 170% 的吞吐量提高（依赖于流量）。认知无线电网络的吞吐量在认知无线电网络的基站的低 SINR 条件下（$T_l=3$）也如图 8-35 所示。这个图中显示了所提出的方案有一个 14% 到 160% 的吞吐量提高。

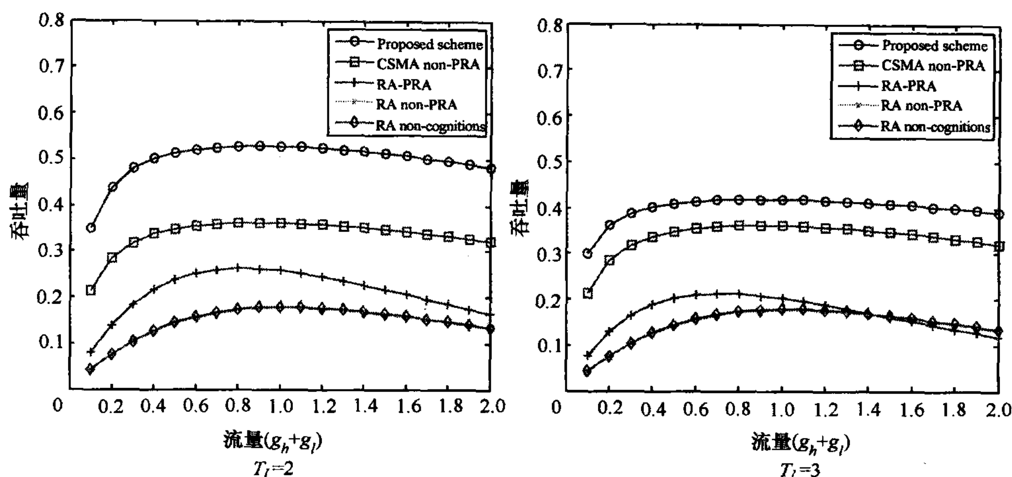


图 8-35　当主系统存在时认知无线电网络的吞吐量

无需位置信息使用空域的无线电资源的 CSMA-ST 可被作为认知无线电网络的多址接入协议，并可带有跨层功率-速率控制机制，在以预定的 PS 截断概率到 PS 的干扰限制条件下，最优化地和并发地跟踪认知无线电网络的吞吐量和总吞吐量。这奠定了关于认知无线电网络的多址接入协议设计的一个重要启示的基础：载波监听和空间复用并发传输对于有效的认知无线电网络中多址接入协议是最根本的。数值结果显示，干扰和信道条件 T_l 调整到 $2T_h$，带有 CSMA-ST 的跨层功率控制机制的认知无线电网络达到吞吐量的 0.53，当 1）无空间复用能力的多址接入方案仅达到吞吐量的 0.36；2）无载波监听或空间复用能力的多址接入方案仅达到吞吐量的 0.18 到 0.26。这个跨层的 CSMA-ST 提升了吞吐量，即使认知无线电网络的隐藏终端问题和 CSMA-ST 的稳定性还是两个未解决的问题。

参 考 文 献

[1] M.A. Marsan, D. Roffinella, 'Multichannel Local Area Network Protocols', *IEEE J. Sel. Areas Commun.*, **1**,1983, 855–897.

[2] H. Okada, Y. Nomura, Y. Nakanishi, 'Multichannel CSMA/CD Method in Broadband-bus Local Area Networks', *Proc. Globecom*, 1984, 642–647.

[3] K.C. Chua, 'Performance Analysis of Multichannel CSMA/CD Network with Noisy Channels', *IEEE International Conference*, **3**, June 1991, pp. 1367–1371.

[4] A.R. Rhaghavan, C.W. Baum, 'Multichannel channel-access schemes for distributed direct-sequence networks', in *Proceedings of the IEEE Military Communications Conference* (MILCOM '98), 1998.

[5] Y.-C. Tseng, S.-L. Wu, C.-Y. Lin, J.-P. Sheu. 'A Multi-Channel Mac Protocol with Power Control for Multi-hop Mobile Ad Hoc Networks', in *Proceedings of 21st International Conference on Distributed Computing Systems Workshops*, April 2001, pp. 419–424.

[6] N. Jain, S.R. Das, A. Nasipuri, 'A Multichannel CSMA MAC Protocol with Receiver-based Channel Selection for Multihop Wireless Networks', *Proceeding of IEEE Vehicular Technology Conference*, 2001.

[7] S.-L. Wu, Y. Lin, Y.-C. Tseng, J.-P. Sheu, 'A New Multi-Channel MAC Protocol with On-Demand Channel Assignment for Mobile Ad Hoc Networks', *Proc. Int'l Symp. Parallel Architectures, Algorithms and Networks* (ISPAN '00), Dec. 2000, pp. 232.

[8] S.-L. Wu, C.-Y. Lin, Y.-C. Tseng, C.-Y. Lin, J.-P. Sheu, 'A Multi-Channel MAC Protocol with Power Control for Multi-Hop Mobile Ad Hoc Networks', *The Computer Journal*, **45**(1), 2002, 101–110.

[9] W.-C. Hung, K.L.E. Law, A. Leon-Garcia, 'A Dynamic Multi-Channel MAC for Ad Hoc LAN', *Proc. 21st Biennial Symp. Comm.*, June 2002, pp. 31–35.

[10] A. Tzamaloukas, J.J. Garcia-Luna-Aceves, 'Channel-Hopping Multiple Access', *Proc. IEEE Int'l Conf. Comm.* (ICC '00), June 2000.

[11] A. Tzamaloukas, J. Garcia-Luna-Aceves, 'Channel-Hopping Multiple Access with Packet Trains for Ad Hoc Networks', *Proc. IEEE Device Multimedia Comm.* (MoMuC '00), October 2000.

[12] J. Chen, S. Sheu, C. Yang, 'A New Multichannel Access Protocol for IEEE 802.11 Ad Hoc Wireless LANs', *Proc. 14th IEEE Int'l Symp. Personal, Indoor and Mobile Radio Comm.* (PIMRC '03), 3, September 2003, 2291–2296.

[13] P. Kyasanur, N.H. Vaidya, 'Routing and Link-layer Protocols for Multichannel Multi-Interface Ad-hoc Wireless Networks', *ACM MC2R*, **10**(1), 2006, 31–43.

[14] P. Bahl, R. Chandra, J. Dunagan, 'SSCH: Slotted Seeded Channel Hopping for Capacity Improvement in IEEE 802.11 Ad Hoc Wireless Networks', *Proc. ACM MobiCom*, September 2004

[15] H.W. So, J. Walrand, J. Mo, 'McMAC: A Multi-Channel MAC Proposal for Ad Hoc Wireless Networks,' *Proc. IEEE Wireless Comm. and Networking Conf.* (WCNC '07), March 2007.

[16] J. Mitola III, G.Q. Maguire, 'Cognitive Radio: Making Software Radios More Personal', *IEEE Personal Communications*, August 1999, 13–18.

[17] J. Mitola III, *Cognitive Radio Architecture*, John Wiley & Sons, Inc. New Jersey, 2006.

[18] K.C. Chen, 'Medium Access Control of Wireless LANs for Mobile Computing', *IEEE Network*, September/October 1994, pp. 50–63.

[19] Z. Tang, J. Garcia-Luna-Aceves, 'Hop Reservation Multiple Access (HRMA) for Multichannel Packet Radio Networks', *Proc. Seventh IEEE Int'l Conf. Computer Comm. and Networks.* (IC3N '98), October 1998.

[20] J. So, N. Vaidya, 'Multi-Channel MAC for Ad Hoc Networks: Handling Multi-Channel Hidden Terminals Using a Single Transceiver', *Proc. ACM MobiHoc*, May 2004.

[21] D. Bertsekas, R.G. Gallager, *Data Networks*, Prentice Hall, 1993.

[22] Y.K. Sun, K.C. Chen, D.C Twu, 'Generalized Tree Multiple Access Protocol for Wireless Communications', *Proceeding IEEE PIMRC*, Helsinki, 1997.

[23] K.C. Chen, C.H. Lee, 'Group Randomly Address Polling for Multicell Wireless Networks', *Proceeding IEEE International Conference on Communications*, 1994.

[24] S. Srinivasa, S.A. Jafar 'The Throughput Potential of Cognitive Radio: A Theoretical Perspective', *IEEE Communications Magazine*, 2007.

[25] J. Mitola III, G.Q. Maguire, 'Cognitive Radio: Making Software Radios More Personal', *IEEE Personal Communications*, August 1999, 13–18.

[26] J. Mo, H.W. So, J. Walrand, 'Comparison of Multichannel MAC Protocols', *IEEE Trans. on Mobile Computing*, **7**(1),

2008, 50–65.

[27] S.Y. Lien, C.C. Tseng, K.C. Chen, 'Carrier Sensing Based Multiple Access Protocols for Cognitive Radio Network,' *IEEE International Conference on Communications (ICC)*, Beijing, 2008.

[28] A.J. Goldsmith, S.-G. Chua, 'Variable-Rate Variable-Power MQAM for Fading Channels', *IEEE Trans. Commun.*, **45**(10), 1997, 1218–1230.

[29] A. J. Goldsmith, P. P. Varaiya, 'Capacity of Fading Channels with Channel Side Information', *IEEE Trans. Inform. Theory*, **43**(6), 1997, 1986–1992.

[30] C.-C. Tseng, K.-C. Chen, 'Layerless Design of A Power-Efficient Clustering Algorithm for Wireless Ad Hoc Networks under Fading', *Journal of Wireless Personal Communications*, **44**(1), 2008, 3–26.

[31] C.C. Tseng, S.Y. Lien, K.C. Chen, R. Prasad, 'On the Rate-distance Adaptability of Slotted Aloha,' *IEEE International Conference on Communications (ICC)*, Beijing, 2008.

[32] Q. Zhao, L. Tong, A. Awami, Y. Chen, 'Decentralized Cognitive MAC for Opportunistic Spectrum Access in Ad Hoc Networks: A POMDP Framework', *IEEE Journal on Selected Areas in Communications*, **25**(3), April 2007.

[33] H. Kim, K.G. Shin, 'Efficient Discovery of Spectrum Opportunities with MAC-Layer Sensing in Cognitive Radio Networks', *IEEE Tr. on Mobile Computing*, **7**(5), May 2008.

[34] L.C. Wang, A. Chen, 'Effects of Location Awareness on Concurrent Transmissions for Cognitive Ad Hoc Networks Overlaying Infrastructure-based Systems', to appear in *IEEE Trans. Mobile Computing*, 2008.

[35] L. Lai, H. El Gamal, H. Jiang, H.V. Poor, 'Optimal Medium Access Protocols for Cognitive Radio Networks', submitted for publication.

[36] R.G. Gallager, 'A Perspective on Multiaccess Channels', *IEEE Tr. on Information Theory*, **31**(2), May 1985, pp. 124–142.

第 9 章　网络层的设计

当认知无线电（CR）成功建立了伺机传输链路后，认知无线网络的核心功能将依赖于网络层的设计，特别是路由的设计。而许多其他设计问题，如流量控制、网络无线电资源管理和网络移动性管理都是基于路由设计。一个 CRN 节点可以被看作一个具有动态频谱接入能力，同时具有可编程多无线电功能的节点。本章将构建节点网络层的功能。换言之，一个 CR 节点通过多无线电系统寻找并利用频谱空洞以自组织的方式来传输数据包。因此，显而易见的是在 CRN 中路由问题是网络层功能的核心问题。

CRN 的网络层在给任何网络中的数据包/数据流提供路由之前，首先要具备网络层中的联合功能，也就是说一个 CR 节点要能成功接入整个 CRN（包括主用户）。原则上，CR 在检测到可能的传输机会（也就是说频谱空洞）后必须完成联合，然后通过物理层传输和媒体接入控制执行动态频谱接入（DSA），将数据包从 CR 发射机传输到接收机。这里，CR 接收机可以是一个 CR 或是一个在主系统（PS）中的节点。除了一般的联合（或注册）到一个网络/系统（通常是 PS）外，难点在于要求一个 CR 在一个很短的可用时间窗内快速联合到另一个 CRN 中的节点（可能是另一个 CR 或者是在 PS 或多无线电系统中的一个节点），这一问题我们将在第 10 章中讨论。本章我们着重讨论路由以及其他与路由、自组织相关的问题。

9.1　移动 Ad hoc 网络中的路由

在讨论 CRN 的路由之前，我们首先回顾在移动 Ad hoc 网络和无线传感器网络中已经被研究过的路由问题，因为它们和 CRN 相似。然而，CRN 总的来说属于异构无线网络（具有部分有线网络），这是与 Ad hoc 网络或传感器网络的基本不同点。路由在这种无线异构网络中基本是一个不可解的问题。为了在 CRN 中给路由开发出一种可行的机制，我们将研究在移动 Ad hoc 网络（MANET）中的路由，然后开发出适合 CRN 结构和网络特性的新路由算法。

9.1.1　移动 Ad hoc 网络中的路由

我们可以将移动 Ad hoc 网络（MANET）看作一种多跳分组无线电网络（mh-PRN）。多跳分组无线电网络和移动 Ad hoc 网络中的路由已经被研究了许

多年。MANET 可以看作一群通过没有基础结构的无线链路通信的移动节点。一般假设在有限的路由资源情况下，由于相比起网络的覆盖范围，节点之间的无线电传输距离比较短，MANET 依赖于多跳型的分组传输，并且每个节点本身充当路由器。

基于链路状态或者是距离矢量算法，传统的路由协议的目标是识别出 MANET 中每个节点的最优路由。MANET 中常遇到拓扑结构变化的情况，这表现在周期性更新的传播路径。为了更新并维护，路由信息需要耗费极大的带宽，这是不可行的。对于基于 IP 的 MANET，路由协议大致根据协议是否是连续更新路由表还是按需反应，可以分为先应式和反应式路由。

先应式路由协议，又称为表驱动路由协议，不断地判断网络的连通性和现存可用的发送分组的路由。这种路由协议，由于链路的剧烈变化，明显不适用于像 CRN 这样频繁重配置的移动网络。反应式路由协议，又称为按需路由协议，仅在需要时才查找路由（这就是按需的意思）。动态源路由（DSR）和 Ad hoc 按需距离矢量（AODV）是两种典型的反应式路由协议。在需要路由时，反应式路由协议才调用某种全局搜索方法，如洪泛。这种协议的代价是确定路由的延迟，但却能反应出最新的网络拓扑结构（也就是可用链路）。

9.1.2　CRN 中的路由特点

CRN 和无线 Ad hoc（或传感器）网络中的路由的主要不同点及其难点总结如下：

■ 链路可用性：CRN 链路在 PS 的空闲期才可用，因此在成功的频谱检测之后可以有效地得到动态频谱接入的机会。这导致 CRN 中的链路，特别是那些所涉及的 CR 在作为发射机或接收机时，总体来说是随机可用的，这使得即使所有节点是静止的，CRN 的拓扑仍然是随机的，更不用说具有移动性的 CRN 了。虽然无线 Ad hoc 网络和无线传感器网络具有相似的地方，但是在 CRN 中的链路变化要快得多，这是因为链路的可用时间只是业务数据分组和控制信号分组的两次到达时间之间的一小部分。这就是说，在 CRN 中链路的可用周期是在毫秒量级，而不像在无线网络以秒、分、小时，甚至是以天来计算的。

■ 单向链路：典型的无线网络有双向链路，因为无线电通信是半双工的。而由于非对称的发射功率或者不同的接收机有不同的干扰强度，典型的无线 Ad-hoc和传感网络可能是单向链路。我们可能把单向链路在无线网络中看作不常见的。然而，在 CRN 中，更可能是单向链路，这是由于一个 CR 节点可能只有一次机会来完成一次发送时间期限，而且不能保证有机会能从其他方向发送过来。另外一种可能的情况是一个 CR 节点想要透过现存的 PS（协作式）来中继数据包，但其他方向传输可能是不允许的，反之亦然。一般而言，一个存在认知

无线电节点的链路很可能是单向的。这是 CRN 有别于其他无线网络的地方，特别是在讨论到网络层功能时。

■ 异构无线网络：与传统的无线 Ad hoc 或传感器网络不同的是，CRN 一般是由异构无线网络所组成的（共存的主系统和 CR 节点组成的 Ad hoc 网络）。通常在这样的异构无线网络中，路由要求系统间的切换。然而，CR 链路可能在很短的时间内是可用的，成功地联网依赖于在这样的异构无线网络中的协作中继。从网络安全的角度来考虑，在无线网络中建立 CRN 以损失安全性为代价来提高频谱利用率是存在争议的，因为一个认知无线电节点在很短机会窗内没有足够的时间来得到安全认证。显然为了 CRN 的路由，需要在异构无线网络和认知无线电节点的操作中需要折中。为了确保一条 CRN 链路可以用来实现网络层的功能，我们可能需要回到硬件的操作。假设精灵观测 PS 和 CR 在 CRN 的运行，CR 必须使用频谱空洞窗来完成数据包的传输。假设这一个频谱窗的长度为 T_{window}，显然有

$$T_{\text{window}} \geqslant T_{\text{sense}} + T_{\text{CR-Transmission}} + T_{\text{ramp-up}} + T_{\text{ramp-down}} \tag{9-1}$$

式中，T_{sense}表示保证 CR 传输和获得相关通信参数的最短检测时间，$T_{\text{CR-Transmission}}$表示 CR 数据分组的传输期，而 $T_{\text{ramp-up/down}}$表示传输斜坡（上升、下降）期这个关系式不考虑发送接收双方的传输迟延和处理迟延，这可以看作是上升或下降斜坡持续时间的一部分。频谱空洞的最大持续时间（可用期）可以看作信标信号的持续时间。

显然，我们必须在 CRN 中给链路可用性建立数学模型。因为链路可用或不可用于伺机传输，考虑到链路可用性变化的定时，我们可以采用连续隐马尔可夫链，并且可以从频谱测量数据中统计出该连续马尔可夫链的具体变化速率。图 9-1b 说明一个具有固定定时（比如，PS 的信标信号的时间间隔）的两状态马尔可夫链。其中 A 表示可用链路，N 表示不可用链路。在 CRN 中的节点 X 和 Y 的一条链路可以定义成 X－>Y 和 Y－>X 的两条单向链路。利用简单的两状态嵌入式马尔可夫链就可以对 CRN 的网络层功能进行总体的研究，并且由此可以在链路可用性和单向链路的挑战下有效设计物理层。

9.1.3 MANET 中的动态源路由

动态源路由（DSR）是一种利用路由发现循环来查找路由的按需 Ad hoc 路由协议。DSR 利用源路由，每个节点在它的路由缓存中维护一张完整的路由拓扑表，如图 9-2 所示。

9.1.3.1 基本路由发现

当一个源节点需要传输数据包到某个目的节点时，在它自己的路由缓存中是没有路由信息的，它会发起一次路由发现的过程。源节点首先会准备一个路由请求包（RREQ），这个数据包含有源节点 ID、目的节点 ID、唯一的一个请求 ID、

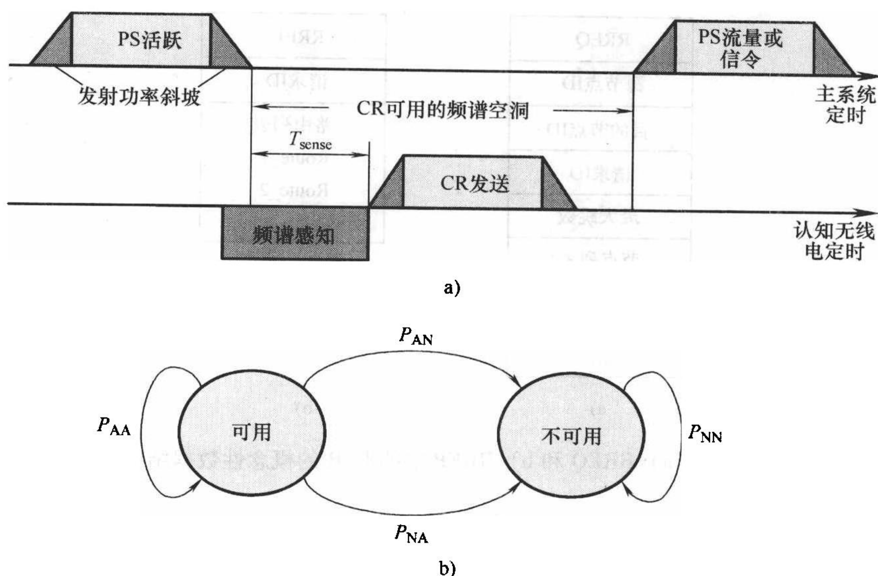


图 9-1　CR 链路可用性模型：
a) CR 发送机会窗　b) 链路可用性隐马尔可夫链状态转移图

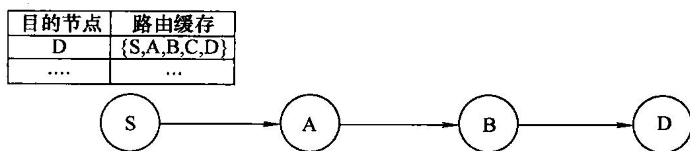


图 9-2　源路由方案的路由缓存

最大跳数的值和一个列出 RREQ 已经经过的所有中间节点的节点列表，如图 9-3a所示。图中描述了节点 S 没有到节点 D 的路由，因此在网络中洪泛 RREQ 包。每个 RREQ 包记录了在节点列表中它所经过的全部节点。

当一个节点收到 RREQ 时，它首先判断自己是否是目的节点。如果不是，为了避免路由环并且节省带宽，它会检查请求 ID 和源节点 ID 对，看它是否收到了一个重复的 RREQ。因为源节点在每一次新的路由发现时，会增大请求 ID，所以一个节点可以通过 RREQ 的源节点 ID 和请求 ID 来区分不同的路由发现。同时，这个节点也可以通过检查自己是否在节点列表中来避免路由环。如果接收到重复的 RREQ 或者它自己已经在节点列表中，这个 RREQ 就会被丢掉。如图 9-4a 所示的例子，F 已经从 E（路径 3）收到一个 RREQ，因此它将丢掉从节点 C（路径 2）收到的 RREQ。最后，如果一个 RREQ 不是重复的也不形成路由环，这个节点就把它的地址添加到节点列表中，并且把这个 RREQ 再广播出去。

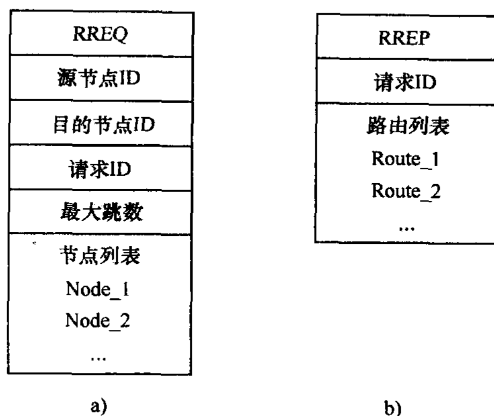


图 9-3 a）RREQ 和 b）RREP 中的 DSR 的概念性数据结构

图 9-4 基本 DSR 路由发现方法

a）S 洪泛 RREQ b）D 有到 S 的路由，并单播 RREP 给 S c）D 没有路由返回 S，洪泛搭载 RREPS 的 RREQ d）S 得到路由，并通过 {S，A，B，D} 回复 RREP 给 D

目的节点的任务只是简单地告诉源节点哪一条路径是可用的路由。因此，它产生一个含有在最新收到的 RREQ 中所累计的路径记录的路由回复信息（RREP），如图 9-4b 所示。然后它检查自己的路由缓存是否有任何到源节点的路由。如果有，它就单播 RREP 回传给源节点，如图 9-4b 所示。如果目的节点没有任何路由记录在它的缓存中，它将开始自己的路由发现，并洪泛携带 RREP 信息的 RREQ 来寻找源节点的路由，如图 9-4c 所示。携带 RREP 信息是为了避免无穷路由发现环。当源节点收到这样一个 RREQ 时，它就知道已找到了到目的节点的路由，并通过这些已发现的路由回复一个 RREP 给目的节点，如图 9-4d 所示。DSR 假设节点之间的链路可以是单向的或是双向的，因此一个路由只表示存在到目的节点的单向链路。目的节点不能直接存储 RREQ 的反向路由。因此，除非路由要求是双向的，否则目的节点不能通过 RREP 的反向路由来回复 RREP 给源节点。这就是为什么一个目的节点在源节点之后可能需要执行第二次路由发现。如果下层要求双向通信，RREP 的反馈必须按照 RREQ 路由的倒序。这个约束条件可以用来检测在已发现的路由中的单向链路。如果路由中的任何链路是单向的，RREP 的传输可能失败，并且源节点将不会再发现这样的路由。如果路由发现失败，源节点将退避一段时间，如果重试次数还没达到对同一个目的节点的最大路由发现次数，源节点将会重试一次。否则，它将通知上层没有路由指示。

9.1.3.2 基本路由维护

路由维护的主要任务是检测拓扑的变化（例如，链路的中断），并通知源节点这些变化。DSR 只在向下一跳节点发送数据包时检测链路中断。各个节点负责确认数据包的成功接收。这可以消除邻近数据包检测的开销（例如，周期性交换 Hello 消息），原因是只有当一个节点需要通过一条链路传递一个数据包时才检测该链路的状态。DSR 使用了三种链路中断检测方法：

■ 链路层确认：如果链路层为数据包的成功接收提供反馈，DSR 可以使用这种方法来确认链路的可用性。这种方法要求两个节点之间的链路是双向的。

■ 主动确认：如果链路层确认不可用，一个节点可以主动侦听它的下一跳的发送，如果下一跳节点发送了数据包，它就确认传输是成功的。这种方法同样需要双向链路。

■ 网络层确认：如果没有可用的方法来判断是否能到达下一跳节点，DSR 在发送数据包前，先在数据包头插入一个确认请求。下一跳节点接收到这样一个请求将产生一个确认，并通过缓存中的可用路由回传给向它请求的节点。由于这个确认可以通过一条多跳路径路由，这种方法可以在单向网络中使用。值得注意的是，收到这样确认的节点不需要再给发送这个确认的节点发送一个确认。

图 9-5 给出了一个反向路由的例子。链路（S，A）、（A，B）和（C，D）

是单向链路，而链路（B，C）是一个双向链路。我们看出每条单向链路有一条反向路由用来链路维护。一个在单向链路中的下游节点可以通过反向路由来反馈一个确认消息给它的上游节点。

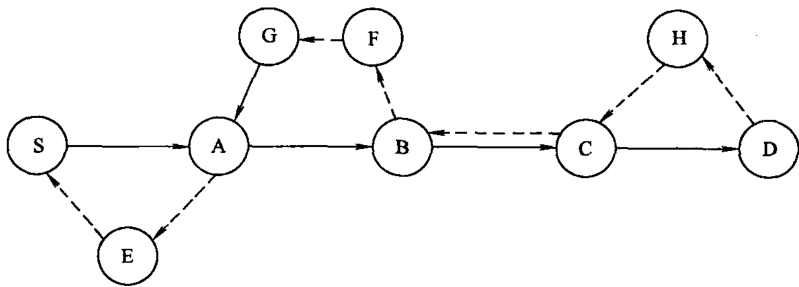


图 9-5 单向链路的反向路由举例

如果一个数据包的发送者检测到一条链路中断了，这样的中断可能是由于没有反向路由或者超出了传输范围，数据包的发送者将返回一个路由错误（RERR）信息给源节点。RRER 包含中断链路。源节点将在它的路由缓存中清除所有包含这条中断链路的路由。除非它的上层要求或者它没有其他路由了，它将不会重新去查找另外新的路由。图 9-6 给出了这两种原因的例子。

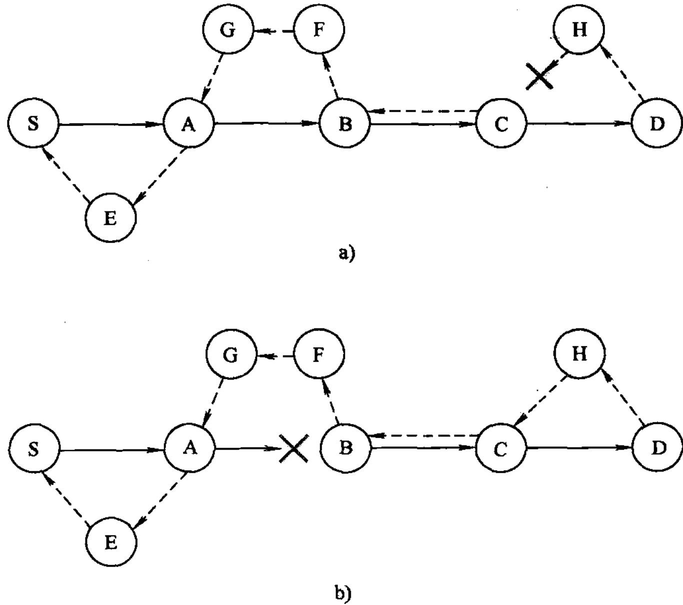


图 9-6 链路断裂的两种原因：

a）由于没有反向路由，链路（C，D）不可用 b）由于超出传输距离，链路（A，B）不可用

9.1.3.3　DSR 的性质

现在我们总结一些关于 DSR 的性质，这可能有助于开发出 CRN 中的路由。

性质 9.1：在 DSR 中，RREP 不一定按照 RREQ 所通过的反向路径传输。

证明：由于 DSR 使用源路由，每一个节点收到一个数据包后，只需要按照附在数据包中的源路由信息传递出去。如果节点本身只是中继数据包，那么它不需要维护路由表。目的节点只需要通过发送 RREP 来通知收到路由的源节点。RREP 所用的路径可能与 RREQ 的反向路径不完全一样。

性质 9.2：在 DSR 中，如果每一个中间节点只重新广播所接收到的属于同一次路由发现的 RREQ 中的第一个 RREQ，那么将会形成路由环。

证明：任何 RREQ 是通过它的源 ID 和请求 ID 来识别的。如果一个节点收到两个或更多个具有相同源 ID 和请求 ID 的 RREQ，那么它可以确认接下来的 RREQ 是重复的。只有当某个 RREQ 不止一次通过同一个节点才会形成路由环。因为节点可以区分重复的 RREQ，并且只重新广播它所收到的第一个 RREQ，所以不会形成路由环。

性质 9.3：如果没有重复的 RREQ 被重新广播，那么当第一跳的节点是不一样时，这两条路由是不相交的。

证明：假设有两个 RREQ，它们的第一跳是被同一个节点接收。图 9-7a 中的路径 2 和路径 3 描述了一个这种情况的例子。假设这两条路径的第一个交汇点就是这个点（例如，节点 F）。因为只有一个 RREQ 可以被广播，其他接收到的 RREQ 将会被丢弃，而且相对应的路径不能用来继续传输到达目的节点。因此两条路径之间不存在交汇点，并且没有公共节点。

性质 9.4：如果没有重复的 RREQ 会被重新广播，那么两条路径只能分开一次，并且没有交汇点。

证明：如果两条路径有相同的第一跳节点，这表示它们在某个节点处分开，否则它们就是相同的路径。如果我们把从分离节点后面的子路径看作通往目的节点的两条具有不同第一跳节点的路径，根据性质 9.3，它们将不会再次相交。因此两条路径一旦分离将不会再次相交。

从性质 9.3 和性质 9.4 中，我们可以知道 DSR 中的路径彼此不会相交，这表示源节点有较少的机会发现非分离路径。如图 9-7a 所示，一个非重复的 RREQ 的传递消除了路径 2。如果目的节点只应答那些第一跳节点不同的 RREQ，DSR 只能发现节点不相交的路径。在 DSR 中，非重复的 RREQ，重新广播和检查节点表可以避免路由环。然而，如果我们放松对重新广播重复的 RREQ 的限制，并且用检查节点表的方法来避免路由环的产生，那么一个源节点可能找到更多到达目的节点的非分离路径，如图 9-7b 所示。

寻找非分离路径可以给源节点和中间节点提供更多的选择来传输数据包，而

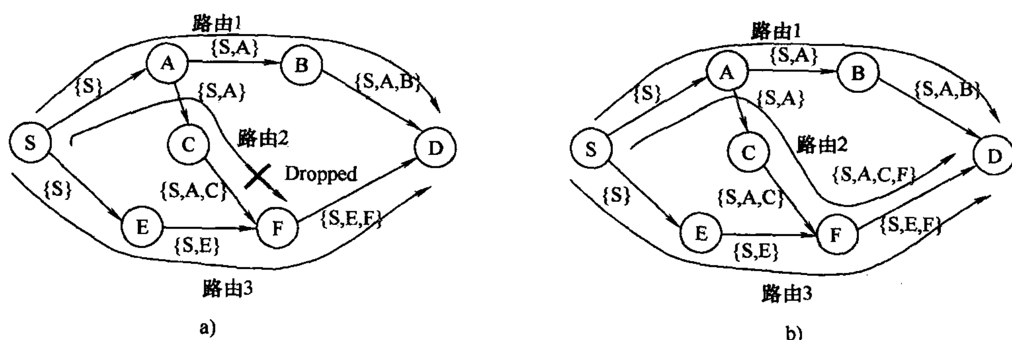


图 9-7 不同的路由发现方法找到不同的路由

a）非重复 RREQ 发送法 b）节点列表检查法

且由此可以增加分组传输率。例如，在图 9-8b 中，各个中间节点有更多的选择来传输数据包给目的节点 D。

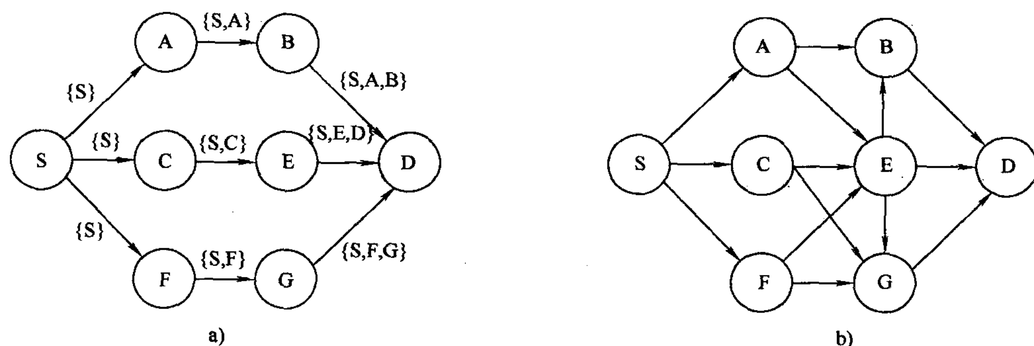


图 9-8 数据包转发中非分离路由的优点

a）节点分离路由发现 b）非分离器由分现

性质 9.5：如果允许重新广播重复的 RREQ，但各个节点会检查节点列表，并丢弃那些已经在它的列表中的 RREQ，那么将不会形成路由环，并且可能找到在最大跳数（最长生存时间（TTL））以内的任何路径。

证明：由于一个 RREQ 通过它之前已经访问过的节点时，它的节点列表中会有这个节点的记录，通过检查节点列表，这个节点很容易发现路由环，并且丢弃这个 RREQ。如果没有限制重新广播重复的 RREQ，那么一条没有环的路径可以到达目的节点当且仅当它具有少于最大跳数的长度。这是因为每一个节点只通过检测节点列表和 TTL 域来判断是否存在路由环和长路径。由于没有路由环，因此可以不超出 TTL 找到所有路径。

性质 9.6：如果使用节点列表法，除了在第一个路径分离点之前的节点，有

k 条不同路径通过的一个节点需要重新广播重复的 RREQk 次。

证明：因为这 k 条不同的路径没有形成环，并且有最大的 TTL，它们相对应的 RREQ 一定能到达目的节点。任何时候，k 个 RREQ 中的一个通过某节点，该节点需要将它重新广播一次。因此一共有 k 次广播。

上述两种 RREQ 转发方法是两种极端的方法，非重复 RREQ 转发可以发现最少的路径，而节点列表检查法可以查找出所有长度小于最大跳数的路径。一种控制发现路由数量的方法是限制各个节点 RREQ 的重新广播的次数。

9.1.4　Ad hoc 按需距离矢量（AODV）

AODV 是一种按需 Ad hoc 路由协议，并且通过路由发现循环来发现路由。AODV 使用逐跳路由，这表示路由信息被分散在网络中的各个节点。每个节点在它的路由表中只维护它下一跳的节点，如图 9-9 所示。

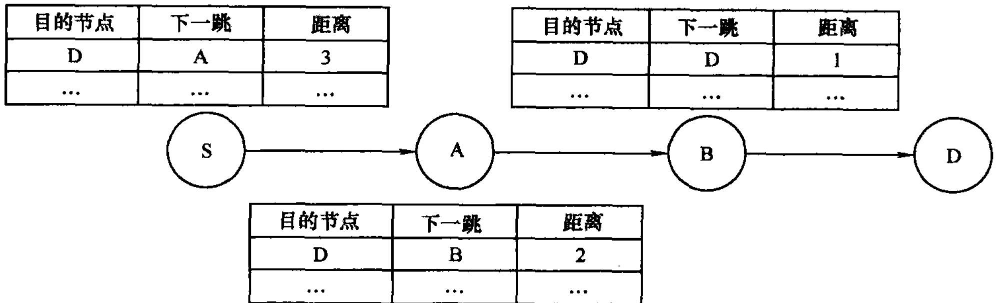


图 9-9　逐跳路由法的路由表

9.1.4.1　基本路由发现

当一个源节点有数据包需要发送到某个目的节点，但在它的路由表中没有到达该节点的路由，它就会发起一次路由发现过程。源节点产生一个路由请求（RREQ）包，这个包含有它自己的 IP 地址和序列号，目的节点的 IP 地址和序列号，还有初始值为零的跳数计数器。

当一个节点收到 RREQ 后，它把跳数计数器的值加 1，并在它的路由表中产生一个反向路由项。反向路由项含有源节点和它所收到的 RREQ 的发送节点。这一项可以用于返回一个路由回复（RREP）到源节点的反向路由中的节点。节点在产生这一项后，它会检查自己是否是目的节点或这一项是否包含到目的节点的路径。在这两种情况下，它会返回一个 RREP，其中含有目的节点的 IP 地址及其最新的序列号和源节点的 IP 地址及其序列号。RREP 也包含了从回复节点到目的节点的距离（跳数）。最后它会根据之前存储的反向路由项来单播 RREP 包。

当一个节点收到 RREP 时，它把跳数计数值加 1 并产生前向路由项，其中含有目的节点和发送回复的节点。这一项可以用来将数据包路由至目的节点。然后节点根据它的反向路由项单播 RREP 给在源节点方向上的下一跳节点。如果源节点收到 RREP，它可以开始向给它发送 RREP 的节点发送数据包。如果源节点收到多个 RREP，它将选择具有到目的节点最小距离（跳数）的路径。

当一个节点单播一个 RREP，它将期待从它所发送 RREP 的节点收到一个路由回复确认（RREP-ACK）。这个操作是为了避免在已发现的路径上使用单向链路。如果没有从哪个节点收到 RREP-ACK 回复，它将哪个节点添加到黑名单中，直到预定的超时值。节点丢弃从它的黑名单上的节点收到的 RREQ。这是为了避免在网络中使用单向链路。

9.1.4.2 基本路由维护

当它们中断时，路由维护机制是用来修复活跃路径。一条活跃路径定义为一条当前正用于传递数据的路径。当一个节点检测到一条链路发生中断时，链路中的上游节点会把所有由于链路中断而无法到达目的节点的路由置为无效。然后它产生一个路由错误（RERR）信息，其中包含了当中失去了目的节点的节点。然后它把 RERR 发送给之前的所有使用该链路到目的节点的节点（因此称为先驱者）。当一个节点收到 RERR 后，它会检查是否真的有到达 RERR 中的那些目的节点的路由。如果有，它就取消所有到达这些目的节点的路由，并且发送这个 RERR 给它的先驱者。最后当源节点收到 RERR 后，它也会使 RERR 中的所有路径失效。如果需要这些目的节点，它将发起另一次路由发现过程。

9.1.4.3 AODV 的性质

我们在此总结一些 Ad hoc 按需距离矢量的性质，这可能有助于 CRN 中的路由（见图 9-10）。

性质 9.7：在 AODV 中，RREP 必须按照 RREQ 所经过的路径的反方向传递。

证明：在 AODV 中，RREQ 到达目的节点后，每个中间节点只知道各自的上游节点和相对应的输入链路，目的节点也是如此。而且，只有下游节点知道输入链路的存在。然而，为了建立路由表，各个中间节点需要知道目的节点方向上的下一跳。因此，让下游节点通知上游节点它有把数据包送到目的节点的路由是十分重要的。在下游节点可以通知它的上游节点之前，它必须先知道自己是否有将数据包送到目的节点的路由，这需要它的下游节点通知它。通过重复这个操作，目的节点必须先通过 RREP 通知它的上游节点。因而 RREP 必须按照 RREQ 所经过的路径的反方向传递。

性质 9.8：基本 AODV 只能发现节点不相交路径。

证明：因为 AODV 要求按照 RREQ 的反向路径传递 RREP，并且各个中间节点只记录一个上游节点和一个下游节点，所以没有两条路径会相交。因而每一次

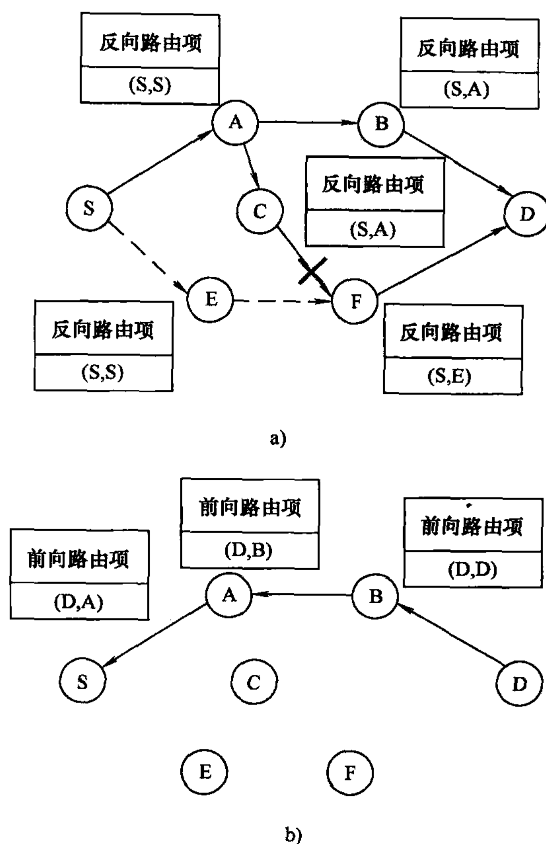


图9-10　基本AODV路由发现方法
a）S洪泛RREQ，每个节点建立反向路由项
b）D通过收到的第一个RREQ的反向路由单播RREP给S

路由发现中的各路径是节点不相交的。

性质9.9：如果没有双向链路提取或单向链路消除机制，那么AODV只能在双向链路网络中使用。

证明：由于必须按照RREQ的反向路径反馈RREP，因而路由发现中的各条链路必须是双向的，否则会因为RREP不能到达源节点而导致路由发现失败。因此，AODV要求链路是双向的。

9.2　CRN中的路由

在列出CRN中路由的三个主要问题后，我们尝试减小第三个问题的难度，然后处理前两个问题以便设计路由算法。

9.2.1 可信 CRN

为了用数学方法解决在前面章节中提到的异构无线网络的问题，我们可以在典型的安全问题上额外引入一套信任机制。请注意，值得观察的是 CRN 中的安全问题依赖于端到端的节点，并且 CRN 中的中间节点（CR 或是 PS 中的节点）可以简单地传递 CR 业务分组（这表示在 CRN 内部的协作中继）。这种协作中继传输数据包的方式可以简化为放大转发（AF）和译码转发（DF），而 CRN 中的中间节点对于中继数据包只有有限的安全措施。压缩转发（CF）协作网络可能由于将中继数据包和自己的业务流混杂在一起而破坏了中间节点的安全性。接下来的协作中继建议使用 AF 或 DF，而不建议 CF。

我们现在可以在 CRN 的操作中，将 CRN 的节点分类及其业务/控制包分为三类：

● 安全：已经经过安全性检查的节点在整个异构无线网络中都被认为是安全的，例如通过公钥基础结构检测。从该节点发出的数据包和信息因为得到安全许可而可以在整个 CRN 中流动。一个属于安全类的节点可以是一个 CR 也可以是共存 PS 中的一个节点。

● 可信任：可信任的安全等级比安全要低。作为一个 CR，一般不能在链路有效的时间窗之内通过几轮的握手协议（即 CR 到 CR，或者 CR 到 PS 节点）来完成一次安全性检测。我们给出一级可信任的安全等级，使得数据包可以在有效的伺机单向链路上传输。在这种情况下，一个 CR 节点在伺机传输中产生数据包，并且 CR 接收机（CR 或 PS 中一个节点）认为这样一个 CR 节点是可信任的，并可以通过合适的路由方法将数据包中继到一个 CR 汇聚节点。请注意，CR 源节点和 CR 汇聚节点将预先用各种方法来完成它们端到端的安全性检测。一个 CR 接收节点应该基于安全性检测和历史更新不断维护一张可信任/安全节点的列表。换言之，任何在 CR 的节点只允许接收从它的安全和可信任邻近节点发送的数据包。这样的一张表是局部性的，并且它的邻近节点数也不大。第 10 章将讨论更新可信任节点列表的方法。

● 诱饵：一个既不安全也不被它的目的接收节点信任的节点被归为“诱饵”类。被归入诱饵类的主要原因是它的不良历史记录，例如散布病毒、浪费无线网络的带宽、攻击无线网络、自私行为等。这样一个诱饵节点实际上丧失了它在实际 CRN 操作中的 CR 功能。

引入一套信任机制的目的很明显是为了给异构无线网络产生一个同构网络的功能环境。因此，尽管 CRN 存在伺机性和剧烈变化的链路可用性，仍然允许协作中继分组。换言之，我们会想方设法通过提供一些激励程序，使得各类无线网络的节点像 CRN 中的节点那样工作（将在第 11 章中讨论）。因而这些节点能高

效地中继从信任的 CR 节点发送来的数据包，继而形成在相同信任等级下跨越不同无线网络的一个大尺度同构多跳 Ad hoc 网络。

我们总结出 CRN 网络层操作的一些关键问题如下：

- CRN 是由 CR 和各类型共存的 PS 的节点所组成，它们可能运行在不同的通信参数下、不同的频带和不同的地理位置。CR 内含的软件无线电能够重新配置实现多系统运行在多个频带上。
- CR 源节点（业务的起始点）和目的节点（业务的终止点）应当在信任等级之外通过引入 CRN 节点来完成双向认证，进而构成它们自己端到端的安全保证。
- 在 CRN 的信任机制下，网络节点被假设只能执行 AF 或 DF 的协作中继。
- 在安全范围内的节点可能拒绝为可信任节点中继数据包，这意味着这样的链路不能用于可信任多跳分组无线电网络的路由。相似的是，在可信任区域中的节点（即 CRN 中的一般节点）可能拒绝诱饵节点的连接请求。
- 任何从一个 CR 源节点发出的数据包，一旦进入 PS 或设备，它就按照 PS 或设备的要求进行操作，这就可以享用多系统或多网络共存带来的好处。例如假设一个 CR 源节点希望通过附近的 WiFi 来中继它的数据包进入因特网的一个网站，“附近”在这里的意思是以无线电可接入为定义的位置。只要允许从 CR 源节点发出的数据包接入 WiFi，这些数据包将被当作 WiFi 数据包来传输。因此要求一个 CR 终端设备能够在多种物理层传输和多种媒体接入控制方法中作转换/重配置。

CRN 的一般操作可以总结为如图 9-11 所示。我们可以使用基础网络（如因特网）作为核心网络架构，同时通过多个无线电接入网络（RAN）提供不同的方式来接入核心网络架构。移动站可以配合某种 RAN 技术。各个 CR 能够按照合适的无线电系统配置自己，以达到为通信/网络传输数据包的目的。无线电接入网、移动站和网络架构可以是任何一种 PS，在图中有几个可能的共存 PS。一个 CR 用户可能是一个 PS 中的移动站。双向链路用双箭头表示，而且所有在 PS 中的链路是双向的。隶属 CR 的动态频谱接入的伺机链路和某些 Ad hoc 链路在图中以单箭头表示。从 CR 源节点出发，有三种不同的协作路径可以传输数据包。同时也有三种协作路径到达作为终点的 CR 汇聚节点。请注意，离开路径 3 和到达路径 3 一般表示之前讨论过的认知无线电中继网络（CRRN）。

我们可以从图 9-11 中清楚地看到，CRN 由 CR 和 PS 组成。CR 动态频谱接入（DSA）使用在一条 CR 链路（或动态频谱接入链路）上的 CR 发射机和接收机中的物理层传输和媒体接入控制。CRN 路由建立在这些 CR 链路和 PS 中的双向链路上。我们再总结一次：

- CR 发射机和接收机通常使用动态频谱接入组成一条 CR 链路。CR 接收机

可以是一个 CR 或是 PS 中的一个节点。

● CR 源节点和目的节点组成像 CRN 中会话层的一段虚拟链路。CR 目的节点可以是一个 CR 也可以是在 PS 中的任何节点。当目的节点是一个 CR 时，我们可以称它为 CR 汇聚节点。

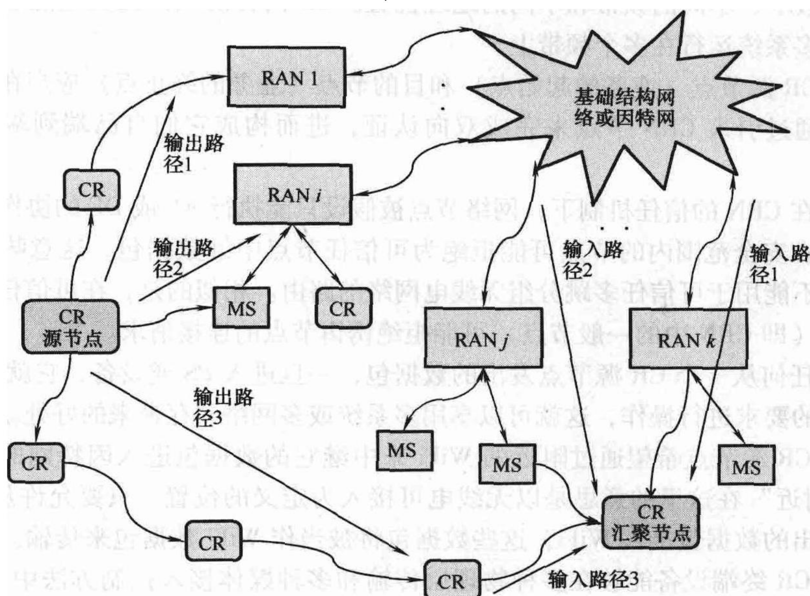


图 9-11 在 CRN 中的路由数据包

9.2.2 CRN 中的动态路由和单向 CR 链路

为了使用前面所介绍的单向 CR 链路和一般情况下 PS 中的双向链路（PS 中的移动站可以通过可能的单向链路组成 Ad hoc）来构造 CRN 的路由，我们可以将 MANET 中的按需路由协议作如下扩展：

● 用二态马尔可夫链对每条 CR 链路建模，各条 CR 链路相互独立。

● 在不知道具体的 PS 情况下，假设 PS 中的所有链路都是双向的，并可以支持我们的路由协议。事实上，如果我们完全遵从 PS 的操作，PS 中的所有行为可以通过对该系统建立排队模型，进而被看作是一条链路。

● 一般 MANET 的路由算法试图孤立各条单向链路，因为它们很可能是非常局部的。然而，单向链路在 CRN 中是不可避免的。幸运的是，我们可以假设从 CR 到 PS 的深度（跳数）在 Δ 跳以内，这是因为它们作用在无线接入到基础结构网络或者接入到纯 Ad hoc。

● CR 链路在某一时刻是单向的，而在下一时刻，这条 CR 链路可能仍然是反

方向上的单向链路，这取决于网络的状态。通过引入信任机制，CRN 可以具有类似带有临时单向链路的 Ad hoc 网络那样的功能。

为了在 CRN 中的路由，我们需要评估 CRN 的一个主要目的：通过更多的协作路径来减小数据流的延迟时间，特别是不使用 CRN 技术，CR 源节点无法将数据包传输到目的节点。同时，有一些我们需要在 CRN 中确定的问题：

● 由于 CR 节点不能干扰 PS，我们应当避免全局或周期性的给任何 CR 节点发公告，而这种公告在 Ad hoc 网络路由中是很常见的。

● 对于一条以 CR 作为发射机的 CR 链路，我们将避免链路上的确认，因为可能没有足够长的伺机时间窗来执行这样的确认。

● 由于和以上两点相同的原因，我们在一般的 Ad hoc 网络路由中不使用 hello 分组。

● CRN 路由必须能检测并尽可能减小任何环或封闭端的可能性，这里的封闭端意思是在一定合理的时间内无法向前传递数据包，而环的意思是传给其他路径的数据包将通过重复的路径返回。

我们假设在 CRN 中的路由考虑局部连通性，这是对 CRN 的操作和路由很实际的考虑，因为 CR 链路只是伺机的。在高度动态变化的特性下，尝试大规模或者全局的优化可能是不可行的。我们的策略是通过高效的伺机 CR 链路来传输数据包到合适的方向/趋势。这完全符合 Ad hoc 网络中的反应式（或按需）路由的思想。

因此，仿照 AODV 路由的原理，我们提出如下 CRN 按需（CRNO）路由。每一个 CR 节点仅在需要时执行路由。路由信息包含以下路由开销信息：

● CR 目的节点 IP；

● CR 源节点 IP；

● 信息 ID（即 msg_id）；

● CR 中继节点 IP（cr_rely_ip）；

● CR 发射机 IP（cr-tx_ip）及其所接收到的数据包/帧的无线电发射类型（cr_tx_type）；

● CR 接收机 IP（crrx_ip）及其所传输的数据包/帧的无线电发射类型（cr_rx_type）；

● 对应的路径的序列号计数器 seq_count（cr_tx_ip，cr_relay_ip，cr_rx_ip），从零开始，经过相同的路径加 1；

● CR 中继节点的计时器，time_counter，从零开始，每一个新的时隙加 1。

当一个新的 CR 节点或 PS 的一个新的移动站进入到网络中，我们可能不能立即获得它的 IP 地址，因此我们为了在表中的操作可以使用一个 ID。

CRNO 路由包括三个阶段的操作：检测阶段、路径发现阶段和表更新阶段

(见图9-12)。

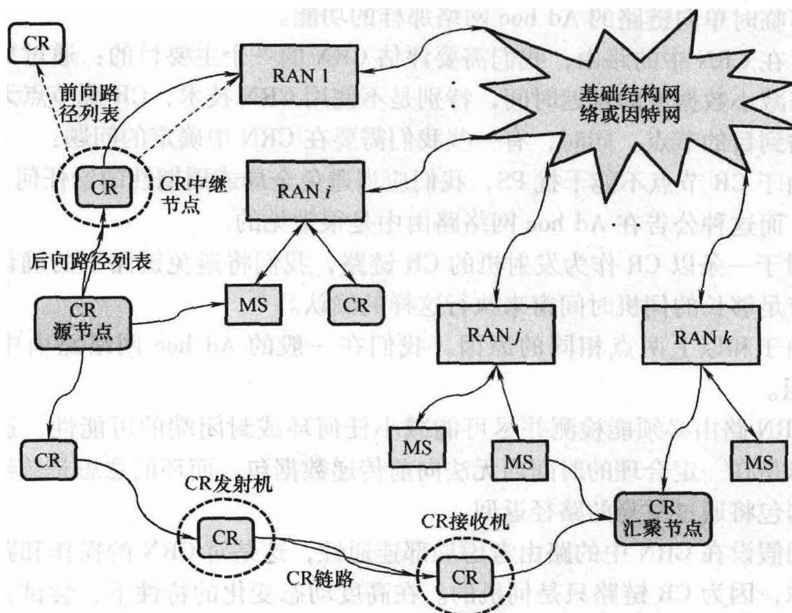


图9-12 CRNO路由

9.2.2.1 检测阶段

CR节点通过监听无线电环境，也就是多共存系统频谱检测（甚至可能是不同的频段），更新它的前向路径表。前向路径表记录了关于潜在的CR接收机、历史记录、它对CR节点的信任度和通信参数的预测，以便调整软件无线电。每一个潜在的CR接收机是通过它之前的传输所得到的IP地址或者通过分配给CR节点的ID来识别。历史记录可以是一个用来根据历史记录和学习过程来指示潜在CR接收机是否是可信任的标识。最后，可以通过频谱检测获得通信系统的参数来调整软件无线电。

9.2.2.2 路径发现阶段

一旦CR节点向目的节点发起一个数据包/帧或从中继节点接收到一个数据包/帧，它将检查后向路径表的任何异常。如果检测没有异常，CR节点从前向路径表另外选择认知无线节点来中继数据包/帧。选择是根据CR链路的可用性和前向路径表。当然，通往PS的链路具有最高的优先级。另一方面，当发生异常时，认知无线中继节点将寻找机会根据后向路径表来否定确认CR发射机。如果可能的话，CR发射节点将尝试重新路由数据包到其他CR中继节点，或者如果没有可用路由的话就进一步回传。

9.2.2.3 表更新阶段

除了选择链路来完成路由，后向路径及与它相关的中继必须作为后向路径表的一部分进行更新。每一条后向路径路由包含的参数有 msg_id、cr_rx_ip、cr_rx_type、cr_tx_ip、cr_tx_type 和 seq_count。

cr_rx_type 和 cr_tx_type 在 CRN 中都是用来指定共存多无线电系统（或重叠无线系统/网络）的运行。

前向路径表中参数的历史记录显然在路由中扮演着一个重要角色。

后向路径表防止环和死端而造成路由灾难。异常的定义是检测到环的存在或是路径的存在。Seq_count 起着判断环是否存在的作用。超时表明为了避免死端而不能中继数据包，这是对更新后向路径表有用的信息。

当否定确认（或从目的节点的肯定确认）不能回溯所有到 CR 源节点的路径时，这可能是由于一些永久的单向链路所造成的；端到端超时可以终止路由，并重新发起一次新的路由。新路由应当避开之前所识别的永久单向链路，并隔离它。

9.3 CRN 的控制

9.3.1 CRN 的流量控制

流量控制在 CRN 中有两种类型。一种可能是 CR 源节点和目的节点之间端到端的流量控制，它会使用典型的基于信用的流量控制，例如漏桶算法。然而，要完全成功地运行诸如 CRN-ODV 或 CRN-DSR 等 CRN 按需路由协议，我们需要其他功能：CRN 中网络层的流量控制。对比传统的计算机网络中的第一类流量控制，CRN 中的流量控制主要是出于损害控制。由于我们不太确定在 AODV 中一定没有死端或环，我们需要检测出这两种情况，并中止 CR 链路中继数据，以保证网络带宽不会被浪费。为了达到这样的目的，需要环检测和死端检测，并配合路由。

进一步地，我们可以观测到如图 9-11 所示的整个 CRN 中的 CR 和 PS 实际上是由如图 9-13 所示的几段所组成的，而数据包通过以下的分段从 CR 源节点路由至目的节点（或 CR 汇聚节点）：

- 上行链路 CRN；
- 一般具有基础结构网络或核心网共存的多无线电 PS（例如因特网），它可能被看作一种在 CRN 中用来快速传输数据包的隧道；
- 下行链路 CRN；
- 认知无线电中继网络（第 6 章介绍的 CRRN）。

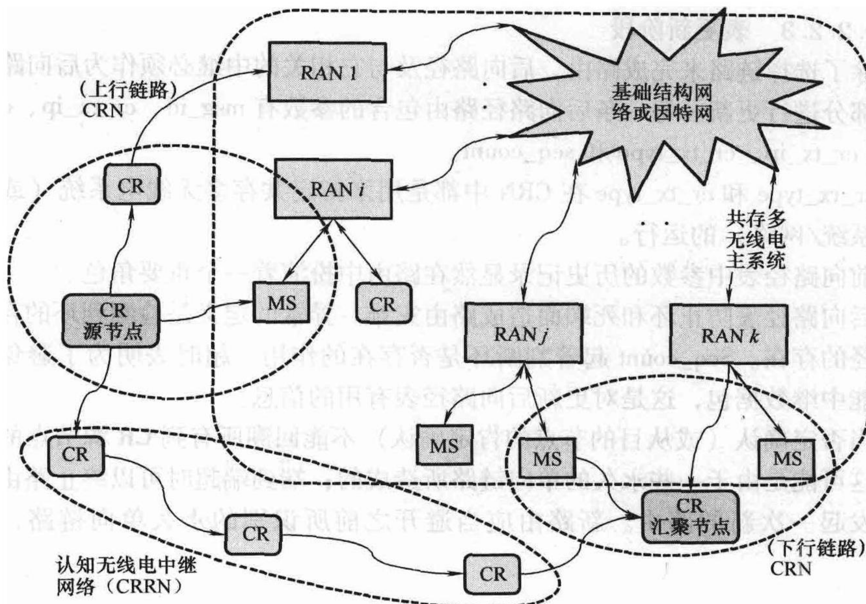


图 9-13 CRN 的分段或分解

CRRN 可以看作是一种由纯 CR 节点组成的特殊 CRN，它唯一的目的就是中继数据包。

业务流量可以分成以下几类：

- CR 源节点 ⟶ 上行链路 CRN ⟶ PS 和基础结构网络 ⟶ 下行链路 CRN ⟶ CR 目的节点；
- CR 源节点 ⟶ CRRN ⟶ CR 目的节点。

因此 CRN 中的路由具有其他基于分段或分解的隐藏的日程表。对于上行链路 CRN，路由试图通过 CR 链路到达 PS。例如，在图 9-13 中，当 CR 中继节点在处理选择前向路径时，它趋向于选择较靠近 PS 的节点，也就是在 RAN1 的节点。然而，对于下行链路 CRN，路由将试图通过伺机 CR 链路离开 PS。当一个在路径发现阶段的 CR 节点主要依据是前向路径表时，参数（或更准确地说是域）历史记录为在节点选择中提供这样的信息起了关键作用。换言之，路由在上行链路 CRN 和下行链路 CRN 并不完全是随机的，是会随着 CRN 内部的动态拓扑而改变方向。这让我们想到蚁群移动，并且参考文献［7，8］利用蚁群路由为开发出表的高效参数/域历史记录更新提供了许多参考。

我们同样注意到 CRN 路由偏向选择能使重叠/共存的多无线电系统更高效的数据包传递方式。这意味着在给一个 CR 中继节点中继数据包时偏向更大范围的 PS（只要它是在可能的选项中），因而这是另外一项 CRN 路由效率的潜在改进。

9.3.2　CRN 中端到端的差错控制

传统概念上的数据包差错控制位于物理层和数据链路层。而差错控制将有利于支持 CRN 的功能。请回顾在 CRN 中的链路是动态可用的，并且它可能不能在一条 CRN 链路上的 CR 发射机和接收机之间执行 ARQ。而且，CRN 路由会尽可能向前传递数据包，这样 CR 汇聚节点可能会接收到同一次发送的数据包的多个副本，而这些同一个数据包的副本由于除了前向差错控制（FEC）之外没有其他的差错保护而可能不正确。物理层的 FEC、CRC 校验和数据链路控制可以保证常规的网络层所要求的极其低的误包率。对于 CRN，不一定存在数据链路控制，并且需要在 CR 源节点和汇聚节点之间的差错控制，同时由于比一般的（无线）网络高得多的代价，将最大限度减小重传次数。我们可以立即借鉴混合自动重传请求（HARQ）的原理来构造 CRN 网络层的差错控制，大大减小 CR 源节点和目的节点之间的业务差错控制。如图 9-11 所示，为了达到在无线网络中可靠的传输数据包的目的，CR 目的节点可能会从 CR 源节点中收到三份或更多的副本，这表示 HARQ 的应用获得更大的路径增益，并且增强了差错控制能力。

CRN 中的 HARQ 的问题是，在 CR 目的节点处，不能确定接收到同一个数据包的副本个数和这些副本的到达时间。

9.3.3　数值实例

一般的 CRN 路由是一种极其复杂的方法。然而，我们可以设计实验来验证我们提出的路由方法。第一个实验是用来论证通过推广在 CR 和 PS 节点中的协作中继，使其能把数据包从 CR 源节点传输到目的节点的 CRN 的可行性。

这个仿真的目标是计算当路由路径是建立在可用信道上的路由延时。路由延迟定义为不考虑其他因素（如发送延时、处理延时等）时，通过这些动态伺机链路而造成的延时。我们按照如图 9-14 所示的拓扑来仿真，并遵从以下假设：

● 只有一个 CR 源节点和一个 CR 目的节点。链路方向如图 9-14 所示。箭头表示链路的方向。虽然在实际情形中存在单向路径，但是在这一步的仿真中，它们没有特别的影响和功能。

● 每一个节点有一张路由表（前向路径表）总结了潜在可用的路径及其接收节点。我们假设频谱检测是可靠的。

● 到目前为止，没有后向路径表，这是因为接收机不发送确认。

● 当开始发送数据时，节点按顺序检查信道的可用性。网络在这一阶段没有信道具有较高的优先级。例如，在图 9-14 中，源节点 31 有 6 个信道，并且在每一轮它总是从 31a 开始扫描，接着 31b、31c，以此类推。

● 每一个信道都具有基于马尔可夫的可用函数。我们选择合适的种子来产

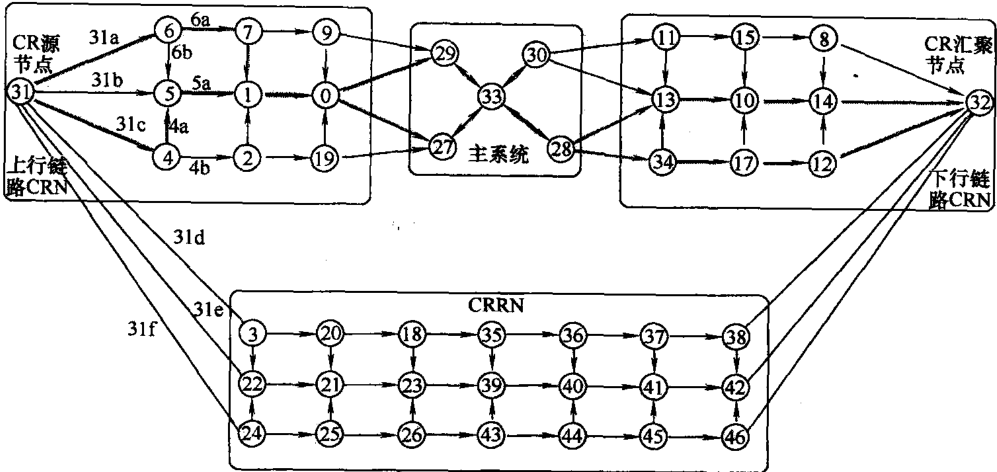


图 9-14 仿真中的 CRN 拓扑结构

生随机数，确保统计的意义。

● 信道传播延迟或在 PS 中的延时可以被忽略；计算延迟只考虑由路由造成的延时。在一个时隙内，一个节点只能扫描一个信道是否可用，延迟计数器是非递减的。

● 在数据包传输期间，零延时表示首先会检查所有信道的可用性，因而不必为了传递数据包而检查第二个信道。

● 仿真重复 10^5次；也就是说发送 10^5个数据包。

图 9-15 总结了图 9-14 给出的 CRN 实例的仿真结果。连通性表示一条链路能用于 CR 传输的概率。根据广泛的研究，PS 的频谱可能只用了 10% ~20% 的占空比，因而可能有参考意义的连通性的值是 90% 和 80%。因此，可以清楚地看到我们提出的路由方案可以在动态可用的单向链路的条件下工作，且在一个表现良好的一般网络拓扑条件下具有可容忍的路由延迟。变量最大重传尝试或生存时间（TTL）表明一个（CR）中继节点上一个数据包的生存时间。当超出这个最大重传尝试或 TTL，中继节点将丢弃该数据包，稍后源节点将通过后向表或端到端控制功能获知包被丢弃。

我们现在考虑一种更灵活的网络拓扑来验证我们提出的 CRN 路由。回顾图 9-13 的 CRN 分解，最一般的路径可以看作是 CR 源节点到 CR 到 PS 隧道到 CR 再到 CR 目的节点。PS 主干在这里起了类似传播延时的隧道效应（假设单位时隙划分时间轴）。对于 CRRN，我们可以简单地消除 PS 隧道效应。对于每一跳，我们同样假设数据包的发送延时是一个单位时隙。现在问题是计算在单向链路条件下从 CR 源节点到 CR 目的节点累计延时（链路可能不可用而导致延时增加）。

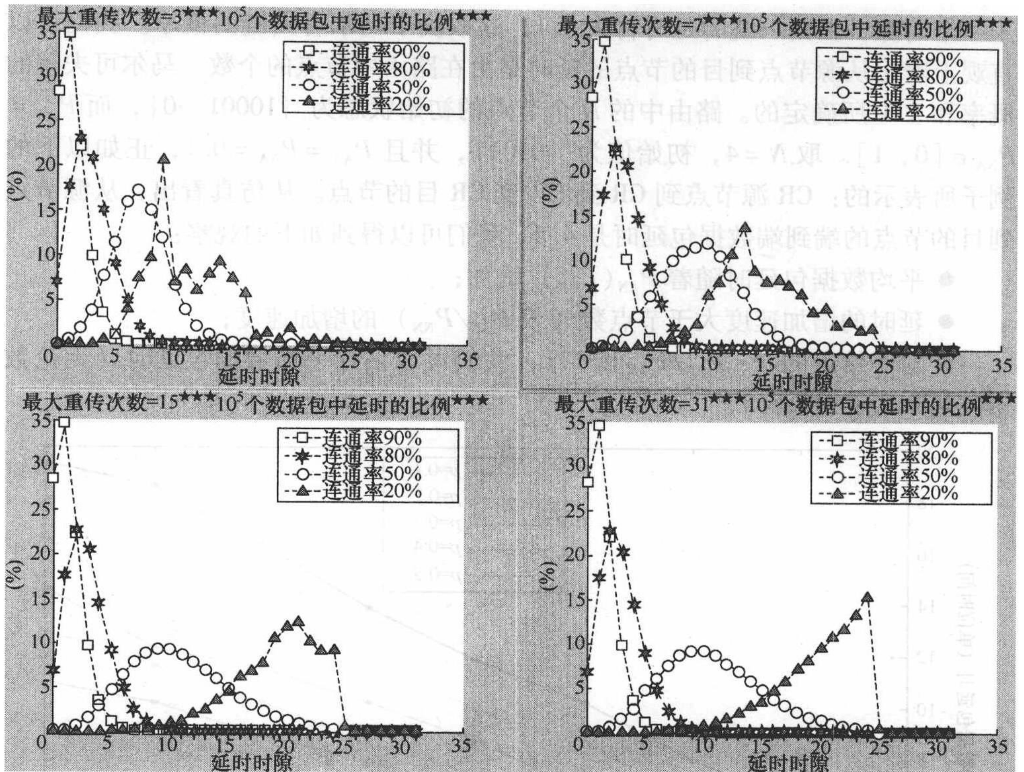


图 9-15　由信道连通性而造成的路由延时的分布，(可供 CR 伺机传输的概率)
x 轴表示路由延时，y 轴表示数据包的比例

首先，我们将研究一维的情况（线性的情况）（见图 9-16)，其中 CR 节点的状态转移可以通过图 9-1b 中的马尔可夫链建模，有 $P_{AA}+P_{AN}=1$ 和 $P_{NA}+P_{NN}=1$。因为不能在单向链路的条件下保证 CR 源节点和 CR 目的节点之间端到端路由，其中的网络拓扑可能变化得很快，所以每一个数据包应当直接从一个节点送到另一个节点。

每一个节点有两个状态，一个是可用状态（状态 A)，另一个是不可用状态（状态 N)。当一个数据包到达某个节点时，它将等待一个或多个时隙，直到节点状态变为 1。例如，如果之前的状态是 0，这表示当前链路不可用，数据包将等待下一个时隙。之后，如果状态值以概率 P_{NA} 变为 1，数据包将根据当前节点的路由表发送到下一跳。如果状态值以概率 P_{NN} 仍然是 0，数据包必须等待下一时隙。最后，如果超过最大时隙，数据包就会被丢弃。

现在，我们假设有 N 个节点在一条从源节点到目的节点的路由路径上。每一个节点是以马尔可夫链状态转移的单向节点。如果链路是可用的，每一跳等效

于一个时隙，否则一跳的延时将大于1，这取决于马尔可夫链的概率。我们可以直观上看出从源节点到目的节点的延时是由在路由中节点的个数、马尔可夫链的概率和TTL所确定的。路由中的N个节点的初始状态为{10001…0}，而$P_{AA}=P_{NA}\in[0,\ 1]$。取$N=4$，初始值为{0101}，并且$P_{AA}=P_{NA}=0.1$，正如以下的例子所表示的：CR源节点到CR到CR到CR目的节点。从仿真看出，从源节点到目的节点的端到端数据包延时是4.5。我们可以得到如下的观察：

- 平均数据包延时随着$P_{AN}(P_{NN})$增加；
- 延时的增加速度大于节点数（或P_{AN}/P_{NN}）的增加速度；
- 对于给定的N、P_{AN}/P_{NN}和TTL，我们可以估计平均数据包延时（无论数据包是到达或被丢弃）。

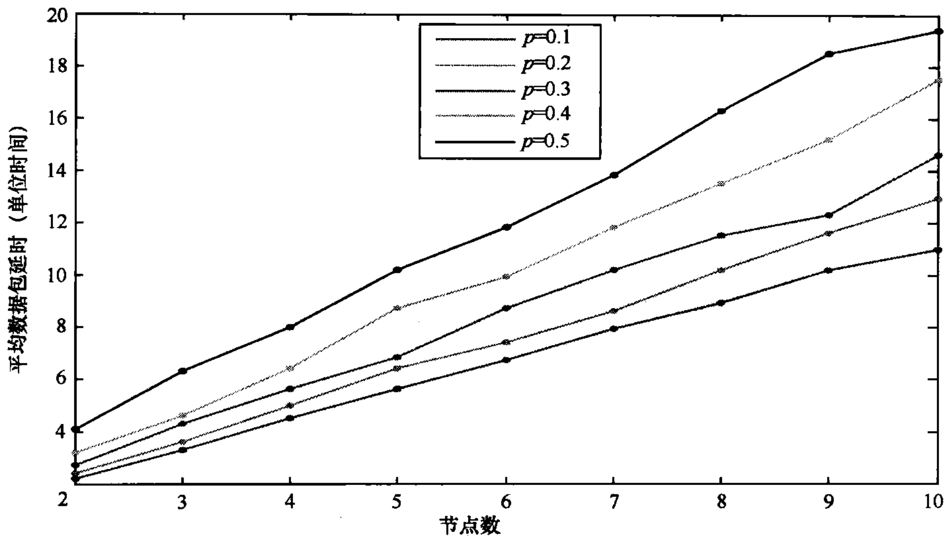


图9-16 线性拓扑的迟时性能

接下来，我们考虑如图9-17所示的在二维场景下的一般网络拓扑。我们的仿真假设50个CR节点随机分布在一个200（单位长度）×200（单位长度）的矩形区域。每一个CR的通信距离是50，也就是说，$N=50$，$[X,\ Y]=200\times200$（单位长度）^{2}，$R=50$（单位长度）。50个CR中任何一个都希望向位于坐标为(100，225)区域外的CR目的节点（CR汇聚节点）传输一个数据包。我们暂时不考虑MS/PS隧道，并且这是一个纯的CRN。稍后可以通过引入这样的隧道来提高网络性能。

基于我们的路由算法，可以用一种贪婪算法建立相应的路由表，最终到达CR目的节点。假设源节点（4）有数据要发送到CR目的节点。它采取的路由是{4→25→8→14→38→17}（见图9-18）。考虑到具有马尔可夫链性质的单向链

路，我们基于以上讨论的一维情况的经验来计算累积端到端数据包延时。如果 $P_{AN}=P_{NN}=0.1$，并且 TTL = 10，我们可以得到平均数据包延时为 6.7。如果 $P_{AN}=P_{NN}=0.4$，最后的端到端延时为 9.9，这小于 TTL。如果 $P_{AN}=P_{NN}=0.5$，从节点 4 发出的数据包将会被丢弃，因为总的延时大于 TTL。在这种情况下，数据包不能成功的传输到 CR 汇聚节点。

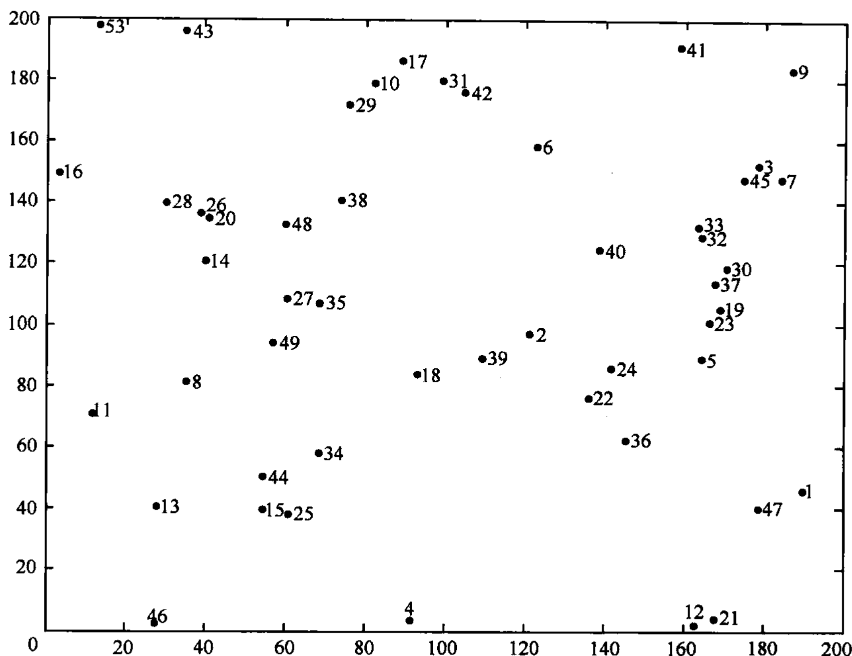


图 9-17　二维随机网络拓扑

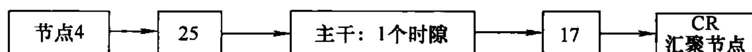


图 9-18　随机产生的 CRN 拓扑的分段

有趣的是，网络拓扑（密度）在数据包传输延时上起了重要作用。如果 CR 增加它的通信范围，网络密度会更大，从而到达 CR 汇聚节点的延时会更小。另一方面，干扰和能量消耗也会随之增加。延时和能量消耗之间需要折中。对于相同的网络拓扑，如果 $R=80$，节点 4 对应的路由是 {4→18→6}，而相应的端到端数据包延时变得更小。最后，如果将一部分 CR 节点替换成 MS/PS，由于在 CR 源节点和目的节点（汇聚节点）之间有了主干网络，端到端数据包延时将进一步减小。用相同的网络拓扑作为例子，但 $R=50$，如果我们将节点 8 到节点 38 替换成主干网络，从节点 4 到 CR 汇聚节点的路由如下，当 $P_{AN}=P_{NN}=0.1$ 时，

总的平均端到端数据包延时为3.3+1=4.3。

9.4 网络断层扫描

正如之前介绍的 CRN 的路由以及相关网络层的功能，我们注意到需要关于潜在可用邻近链路的信息（可能多达几条链路的深度）。进一步地，因为 CRN 很容易变动，并且可以在共存多无线电网络中内部相连于异构网络。所以，端到端的路由似乎不可能，因此网络层的关于本地潜在可用链路的信息尤为重要。诸如反映出 CR 链路可用性的业务活跃量信息对于路由和其他网络层功能是十分重要的。

虽然我们在第 7 章已经介绍过用网络信息检测共存多无线电系统的技术，但是要从 CRN 中的潜在链路获得这些网络层的信息需要不同的方法。直接的方法是利用现有的网络断层扫描技术[10,11]。网络断层扫描原来用于在像因特网那样的大型网络中进行网络监视和推断，在没有额外的节点之间协作的情况下，获得信息。网络断层扫描这个学术名词是由网络推断和医疗断层扫描结合而得到的。通常考虑的网络断层扫描有两种：

- 基于端到端的链路层的参数估计和路径层的业务测量：业务测量通常包括两个节点之间的数据包发送和接收计数，或者两个数据包发送和接收之间的延时。延时是由于传播、排队和路由。路径延时的测量是所有链路延时的总和。一个数据包可能会被丢弃，因此延时一般是随机的。
- 基于链路层业务量测量的发送-接收（目前我们考虑的情况是 CR 源节点和目的节点）路径层的业务量估计：基于数据包计数，目的要估计从某一个节点开始到另一节点的业务量。一个业务矩阵所有起点-终点对的业务的组合本质上是随机的（对于链路层和路径层）。

我们通常假设网络断层扫描中的协作节点会发送一些探测数据包。大多数的网络断层扫描问题继而可以从线性模型开始：

$$\boldsymbol{y}=\boldsymbol{R}\boldsymbol{\theta}+\boldsymbol{\varepsilon} \tag{9-2}$$

式中，$\boldsymbol{y}$ 是观测（或测量）矢量，$\boldsymbol{R}$ 表示路由矩阵，$\boldsymbol{\theta}$ 是一个数据包参数矢量，$\boldsymbol{\varepsilon}$ 表示噪声矢量。对于因特网的断层扫描，$\boldsymbol{R}$ 是一个二值矩阵（意思是矩阵的每个元素是 1 或是 0）。事实上，为了反映网络中时间变化的性质，我们可以引入时间标并重写等式：

$$\boldsymbol{y}_t=\boldsymbol{R}_t\boldsymbol{\theta}_r+\boldsymbol{\varepsilon}_t$$

对于 CRN，由于它的时变本质，这样是更加合理的。另外的方法是引入一维新的未知量，并且我们将把 $\boldsymbol{R}$ 中各个元素看作［0，1］之间的概率，作为 CRN 中的随机矩阵。

我们现在概述网络断层扫描。数据包从一个源节点被送到几个目的节点，如图 9-19a 所示。从源节点到目的节点的路径可能经过一些我们没有或只有有限信息的区域，图中云状区域内，直接探测是不可行的。逻辑链路的拓扑可以由图 9-19b 来表示。

从例子中，我们考虑如图 9-19c 的最简单的树状拓扑。一般来说，在考虑从源节点到目的节点的一条路径时，我们可以给节点标记为 $j=0,1,\cdots,m$，并且有 n 个不同的测量路径通过网络。我们可以定义 r_{ij} 为第 i 条测量路径包含第 j 条链路的概率。对于图 9-19c，我们有

$$\boldsymbol{R}=\begin{pmatrix}1 & 1 & 0\\ 1 & 0 & 1\end{pmatrix}$$

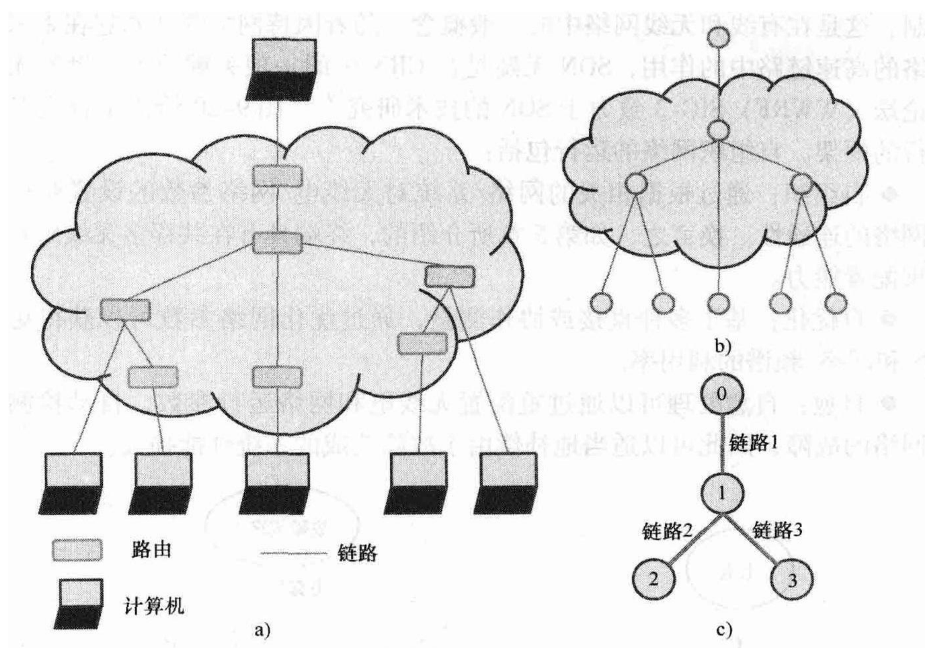


图 9-19　直接探测/测量不能接入云状区域

a）单个源节点多个接收机网络的物理结构　b）逻辑拓扑　c）简单树状拓扑

在这个例子中，矩阵可能不是满秩的。我们可以通过 $\boldsymbol{R}$ 建立链路预测或统计推断出是单播网络还是多播网络，由此得到多种端到端网络特性。请回顾我们在第 7 章讨论的通过主动检测或探测开发 CR 中的无线电资源，这可以认为是一种 CRN 断层扫描，它可以通过统计推理或学习算法开拓一个新的研究方向[15]。我们需要知道更多的这方面的前沿知识以保证 CRN 的成功而流畅的运作。

9.5 在移动通信网中的自组织

为了将来给大量的无线设备提供服务，蜂窝型系统趋向于使用微基站来给小蜂窝提供服务，同时会在有限的网络管理复杂度下，使用更多的中继站来扩大无线电覆盖范围。通过大量的基站和中继站，同时考虑 CRN 的趋势，自组织成为未来移动通信网络的一项重要特征。

9.5.1 自组织网络

自组织网络（SON）系统的概念上的理解是系统不需要任何外在或中心专用控制。这是在有线和无线网络中的一般概念。随着因特网的变迁和它在未来无线网络的高速链路中的作用，SON 无疑是在 CRN 中的一项关键技术。世界无线研究论坛（WWRF）SIG-3 致力于 SON 的技术研究[9]。图 9-20 给出了自组织网络运行的框架。自组织网络的运行包括：

- 自组织：通过根据相关的网络/系统对无线电/网络参数的设置获得终端到网络的连通性。换言之，如第 5 章所介绍的，终端具有在共存多无线电环境中的重配置能力。
- 自优化：基于多种直接或协作测量，通过优化网络参数可以获得更好的 QoS 和网络/频谱的利用率。
- 自愈：自愈处理可以通过重配置无线电和网络运行参数，自动检测并恢复网络的故障，因此可以适当地补偿由于故障造成的系统性能损失。

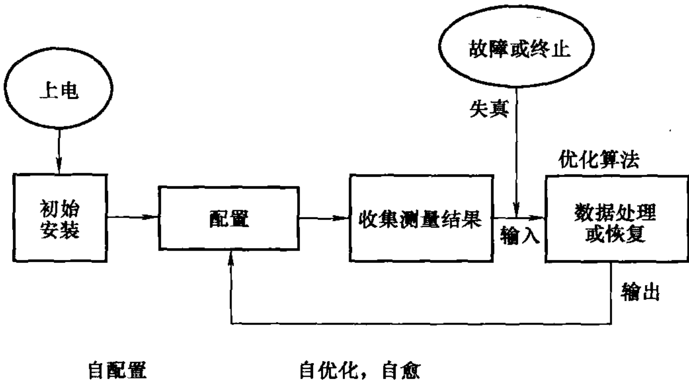


图 9-20 自组织的基本框架

以下几种典型的网络运作情形可以用来进一步解释 SON 是如何工作的：

- 联合：一个新加入无线网络的节点应当执行 IP 地址、检测/联合网络管理

单元、授权、下载节点的无线电软件和网络操作、更新蜂窝信息和关于邻近蜂窝、基站的信息、调整可配置的无线电网络和通信参数等初始化设置和配置过程。

● 节点故障和关闭：性能测量会指示出问题的所在。然后，启动故障检测和恢复 SON 功能的方案，使得网络自愈。

● 运行环境的改变：由于节点的移动性和业务（可能在 CRN 中的任何时刻）的动态变化，自优化和自配置可以用来调整适应任何的改变。

● 异构网络环境：共存多无线电系统需要解决这个问题，并且利用伺机网络这个优点，特别是在 CRN 中。

自配置明显依赖第 2 章所介绍的可重配置无线电收发机和网络功能。自愈依赖于系统/节点的故障检测或报错，这更像是系统感知，除了它更集中于问题的辨识上。最主要的技术问题是如何在 SON 和 CRN 中获得自优化，特别是覆盖范围、切换和共存多无线电网络的无线电资源管理。

9.5.2 自组织协作和认知网络

为了扩大覆盖范围和增强无线电资源的使用，我们可以结合与 SON 协作的概念，组成更健壮的 CRN。我们可以通过内部系统切换[13]或在不同系统之间的无线电资源联合分配达到这个目的。

9.5.2.1 系统间切换

随着未来的无线通信网络要求更大的灵活性、可扩展性、意义识别和在网络架构中的伸缩性，系统间切换可以实现终端接入选择、候选系统和个人网络中的移动管理。IMT-A 候选系统具有结构简单的特点，可以减少它的物理层单元的数量以增强它的逻辑功能（例如，更接近空中接口的移动 RRM 功能）。作为一种移动管理的支持，不同系统之间的通信应当使用通用的接口。多模式终端是 4G 系统考虑的一方面。应当注意的是合法系统（例如，GPRS、UMTS）的配置可能会影响这些终端的操作性。一个 4G 多模终端连接在一个 4G 系统，应当能够对来自一个确定的合法系统用于系统间切换和其他功能的封装信息进行解码。为了能够快速扫频并且在适当的 RAT 中给多模终端进行小区注册，4G 系统应当根据运营商的政策广播从合法系统选择系统的信息。这样的系统信息可以是给某个合法 RAT 分配的频率。

更进一步的是在可行时 4G 系统和全局频谱协调系统之间的互连。全局频谱协调系统运行在一组带外信号，携带关于运营商的频谱协作。这样的互连实现了对于 4G 系统的高级频谱管理。可以通过在 4G-RAN 结构上的部分物理单元中执行一些功能部分来实现支持系统间移动管理的架构。含有 4G 系统的系统间移动管理的一个例子是如图 9-21 所示的一个 WLAN 系统和一个 UMTS 系统。可以通

过引入一个网关（GW）作为系统间通信的定位点，并且这个网关提供了接入因特网的接口与外部路由功能模块进行通信。可以预测在未来，GW将与不同的RAN相连，并且RAN将会被看作整个通信网络的一个要素。一个基站（BS）将可以为使用中的终端（正在发送数据的终端）实现所有无线电相关的功能，并且可以承担管理一个或多个小区中从终端和中继节点（RN）的无线电发射和接收。基站管理中继（如果存在的话）并决定路由，传递数据包给各个中继，并负责中继的流量控制来确保它们可以将数据传递到它们相应的终端。

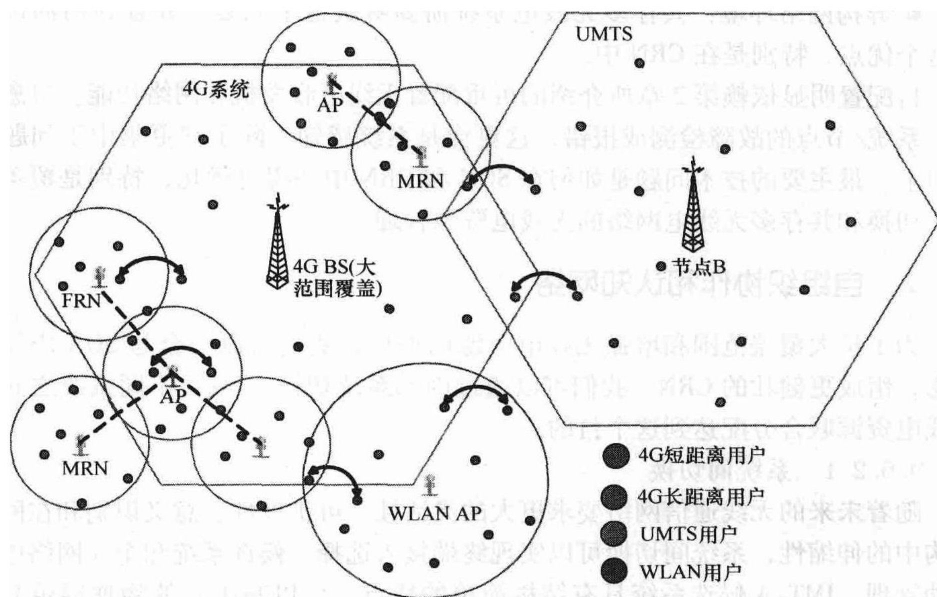


图9-21 包含4G系统的系统间移动管理情形

中继节点拥有中继的能力，通过无线连接到基站、移动终端和其他中继节点。因此它包括转发的功能，可以在无线电接口上进行调度。进一步地，由基站提供的系统信息广播被中继节点中继以扩大系统的覆盖范围。一个基站可以通过多个中继节点和终端进行通信，即多跳通信。

I_G 接口将确保不同系统之间的互连。如果我们假设按照3GPP所提出的将控制和用户平面功能分开，GW控制平面必须通过 I_G 接口支持必要的请求、测量报告和确认。位于RAN架构之外的模块也可以执行与系统间移动性管理相关的互连功能之间的协调工作，这也是通过 I_G 接口进行通信。这与众所周知的RRM（在TR25.881中是CRRM）方法相似，因而作为互连被3GPP采用。一般的RRM功能应当由GW承担，为了简化系统间的互连，同时不会造成太大的信号负荷。

有提议在 4G 场景下的系统间互连是基于紧耦合原理（意思是说负责系统间互连的外部模块将参与各个 RRM 决策）。一个互连架构应当有足够的通用性来涵括任何类型的系统到互连过程，而不需要为各个系统特有的 RRM 机制作出改变。有提议协作的实现是基于如图 9-22 所示的推荐体系结构。一个特有的 RRM（SRRM）是一个负责配合 RAN 和负责系统间 RRM 的外部模块所执行的协作的模块。SRRM 有两种类型的功能和与 RAN 的接口，一种是为了业务监测和报告物理节点的状态，另一种是用于在 RAN 节点的直接启动；基本上它将外部模块 RRM 指令翻译成 RAN。除了内部 RRM 功能，SRRM 在 4G RAN 操作被建议分布在 RAN 节点中（即分别是 GW 和 BS）。

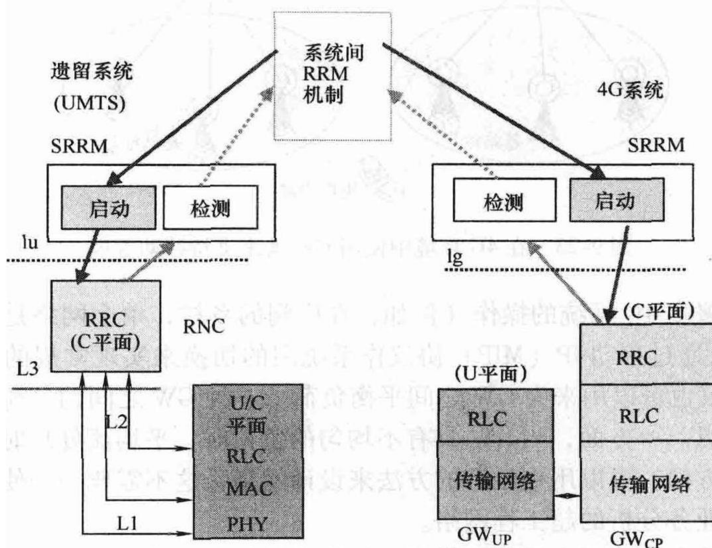


图 9-22 在 4G 环境下的系统间互连的实现

在另一种支持漫游的实现中，不同的互连系统可以通过一个与归属用户服务器（HSS）或归属位置寄存器（HLR）相似的模块相耦合。开放式耦合是另外一种互连的方法，其中没有网络间交换信息的机制。所有的操作由移动终端来协调，这作为系统间切换唯一的中继。这种方法对各种类型的系统都适用；然而，关于新系统的信息和无缝操作的可能性以及可靠的切换是十分有限的，除非在切换中使用先开后合的策略，这一般要求两套无线电收发机。

GW 池的概念[13]（见图 9-23）分离一个唯一的 GW 和与之相连的一定数量的 BS 的物理联系。取而代之的是每个 GW 可能与每个在池范围内的 BS 相关联。因此，默认在一个池范围内所有 GW 是同等平衡的，那么不必引入 GW 之间的切换（HO）。据此，池范围定义为 UT 可能不需要切换 GW 的漫游的范围。GW 对

于一个池范围的容量可以通过增加更多的 GW 来增大。相反，在传统分层结构中，各个 GW 是和一组 BS 相关联，为它们各自所在区域服务，提供 GW 和与之相关的 BS 所覆盖的区域的直接映射。

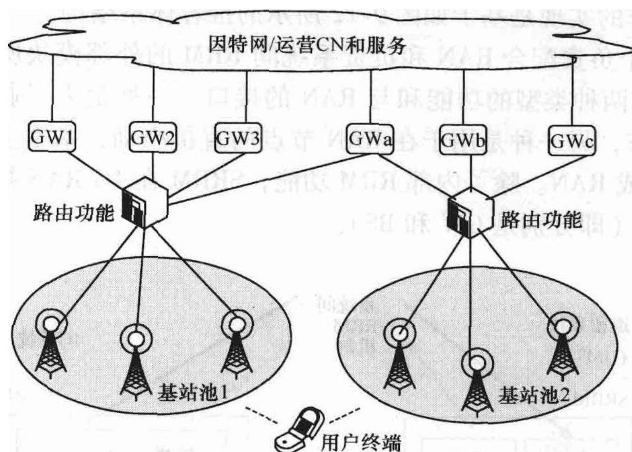


图 9-23 在 4G 环境中使用 GW 池来支持移动管理

可以部署多 4G 系统的操作（例如，在广阔的乡村，单个网络是不足以全部覆盖），并且通过移动 IP（MIP）协议作系统间的切换来实现常规的宏移动性机制。MIP 协议也可以用来为 GW 之间平衡负荷。一种 GW 之间的负荷平衡机制至少有两个优点。一方面，当 GW 具有不均匀的负荷时，平均流处理时间将会明显缩短。另一方面，可以用更经济的方法来设计网络，这不需要一个使用基于传统峰值时刻的业务分析的超工程网络。

网络架构和切换联合设计提供了总体优化性能。它成功引入了一种混合切换机制，这完美地与池范围部署的概念相吻合。GW 使用池范围的关键原因在于负荷平衡以及避免不必要的 IP 切换。这可以使得最大程度地使用系统资源，并且提供系统性能。网络架构可以通过一网关池，也可以是一频带池来实现。正如移动 WiMAX 的系统迁移已经是 3GPP 的一部分，在全 IP 的网络结构下，系统间的切换可以简化为目前能实现的技术和扩展的吉比特无线接入。

9.5.2.2 协作式联合无线电资源分配

一种甚至更为强势的做法是，设立一个高度集成的系统网络架构，它允许在不同的系统中进行联合无线电资源分配，为了逼近无线电资源（指的是频率/子载波、时间、编码、MIMO 系统中的空间/天线）的最大利用率。这是值得考虑的，因为由于不同的应用和需求，有越来越多的无线通信系统会出现。例如，从 1990 年开始 GSM/GPRS 一直为移动电话提供服务，而作为局域网服务 IEEE 802.11 提供了室内中等距离无线接入。蓝牙和 UWB 常用于短距离（~1m）传

输，而作为传感器网络的 Zigbee 则用于一些分布式应用。除了所有上述已有的系统外，还有新出现的无线系统，诸如 3G（CDMA 系列）、HSPA、3G LTE 和 IEEE 802.16 系列。正如我们所知，这些系统都在耗尽带宽，并且使得可用频谱越来越缺乏。把固定的频谱分配给一个指定系统造成频谱的低效，并且在未来将变得越来越严重。

共存/协作通信是解决这个问题的一种方法。在参考文献［14］中，作者为 WiMAX 和 HSDPA 的共存建立了联合无线电资源管理（JRRM）。根据各个系统的实时情况将不同用户的业务分配到这两个系统中的一个。一个更深入的情形是考虑欧洲 WINNER Ⅱ 项目，对于新系统和合法系统，它将移动管理中的 IP 切换和无线电切换融入到共存无线电资源管理。然而，只从网络选择或无线电资源管理考虑共存多系统也许太保守了。这就是为什么我们介绍跨三层无线电资源分配的原因。为了最大化频谱效率，我们尝试将各层设计无线电资源利用的因素集中在系统设计中考虑。同时考虑频谱在小区中的复用、在系统和小区用户之间的带宽分配、天线使用率和对各系统的功率分配。这无疑是一种由上到下的跨层设计，从无线电资源一直到物理层，并且为了实现协作无线电接入和移动管理，自组织网络在各系统间的协调当然是需要的。下面我们将对这些想法一一说明。

我们从一个含有许多小区的网络开始，有几个共存系统在这些小区中。由于大多数无线系统都采用频率复用，我们可以减小频谱重新分配的复杂度，只考虑 N 个小区在一个频率复用模式下。我们可以采用 $N=4$ 的频率复用模式作为一个例子。每个六边形代表一个小区，而不同的颜色表示不同的小区使用的是没有重叠的频带。明显的是，相邻小区会使用不同的频带，以避免同信道干扰。

频谱重植可以看作是按照各个小区的需求分配预定好的频带 B 给 N 个小区，以达到最大网络容量。如果有 X_i 个系统在第 i 个小区中，而第 ε 个系统的容量是 C_i^ε，重植问题可以用式（9-3）来表示，其中 B_i 是分配给小区 i 的频带，并且有 $B_1 \cup B_2 \cup B_3 \cdots B_N = B$。

$$C_T = \max_{B_i} \sum_{i=1}^{N} \sum_{\varepsilon=1}^{X_i} C_i^\varepsilon \tag{9-3}$$

在此我们具体定义各个 C_i^ε。它直接与各个共存系统相关。

我们接下来考虑频谱在各个小区中的分配。在小区 i 中，无论是数据链路、控制链路或是中继链路，所有系统和它们的用户共享相同的带宽 B_i。在此我们重点研究中心的结构；各个系统有一个基站，而用户分布在整个小区内。在这种场景下，我们可以更进一步把无线电资源分为时隙和频带。已经有许多研究报告介绍了如何结合功率将时间-频率资源在一个系统内分配各个用户，使得系统容量最大化，如参考文献[14]。在此我们提出将这些方案扩展到多个共存系统并兼顾上层来考虑分配问题。让我们以 OFDMA 和 MIMO-OFDMA 的共存作为例子；

由于它们的高频谱效率和频谱使用的灵活性，这两种系统已被许多无线系统采用。带宽在这种系统中被分为正交子载波，并且功率可以按照各子载波来分配。如果使用了 MIMO，多天线从另一角度增加了系统容量。可以用式（9-4）表示小区容量的最大化：

$$C_i = \max_{\omega_{t,\alpha,k}^{i,\varepsilon,u} p_{t,\alpha,k}^{i,\varepsilon}} \sum_{\varepsilon=1}^{X_i} \sum_{u=1}^{U_i^\varepsilon} \sum_{t=1}^{T} \sum_{\alpha=1}^{N_T^\varepsilon} \sum_{k=1}^{K_i} \frac{\omega_{t,\alpha,k}^{i,\varepsilon,u} \log(1 + SNR_{t,\alpha,k}^{i,\varepsilon,u}(p_{t,\alpha,k}^{i,\varepsilon}))}{TN_T^\varepsilon K_i} \tag{9-4}$$

在式（9-4）中，我们假设该小区的带宽 B_i 被分为 K_i 个子载波，而我们一次考虑 T 个时隙。对于使用 MIMO 的系统，N_T^ε 表示发射天线的个数：那些除时间和频率外还有更多一维分集的系统。u、t、α 和 k 分别表示用户、时隙、天线和子载波的编号。对于小区 i，我们假设有 U_i^ε 个用户属于系统 ε。式（9-4）中的变量 ω 和 p 分别是分配的子载波数和功率，$\omega_{t,\alpha,k}^{i,\varepsilon,u}=1$ 表示在系统 ε 中，时隙 t 的子载波 k 在天线 α 中被分配给用户 u，否则 $\omega_{t,\alpha,k}^{i,\varepsilon,u}=0$。而 $p_{t,\alpha,k}^{i,\varepsilon}$ 则是分配给该子载波的功率。

我们也不在式（9-4）中具体定义信噪比，因为这也是直接与各个系统的结构与信道情况相关的。但是对于没有移动用户的系统或者处于良好平衡的系统，SNR 可以用式（9-5）表示，其中 X 表示发射的数据，而 H 表示各个子载波的信道增益：

$$SNR_{t,\alpha,k}^{i,\varepsilon,u} = (p_{t,\alpha,k}^{i,\varepsilon} |H_{t,\alpha,k}^{i,\varepsilon,u}|^2 |X_{t,\alpha,k}^{i,\varepsilon}|^2) \Big/ \left(N_0 \frac{B_i}{K_i}\right) \tag{9-5}$$

从这些例子中，我们想到了一些关于共存系统的跨三层资源分配的方法。对于 MIMO-OFDMA/OFDMA 共存的例子，将式（9-4）和式（9-5）代入式（9-3）中得到一种优化方案。如果我们需要考虑如单载波系统那样更为普遍的情况，我们只需要修改式（9-4），然后就可以得到另一种优化方案。

参考文献

[1] J. Mitola, III, *Cognitive Radio Architecture*, John Wiley & Sons, Inc., New Jersey, 2000.

[2] K.-C. Chen, 'Medium Access Control of Wireless LANs for Mobile Computing', *IEEE Network*, September–October, 1994.

[3] D.B. Johnson, D.A. Maltz, 'Dynamic Source Routing in Ad-hoc Wireless Networks', in *Mobile Computing*, T. Imielinski and H. Korth (eds), chapter 5, pp. 153–181, Kluwer Academic Publishers, Boston, MA, 1996.

[4] C.E. Perkins, E.M. Royer, 'Ad-hoc On-demand Distance Vector Routing', *Proceedings of the second IEEE Workshop on Mobile Computing Systems and Applications*, 1999.

[5] A. Boukerche, 'Performance Evaluation of Routing Protocols for Ad-hoc Networks', *Mobile Networks and Applications*, **9**,2004, 333–342.

[6] R. Prakash, 'A Routing Algorithm for Wireless Ad-hoc Networks with Unidirectional Links', *Wireless Networks*, **7**,2001, 617–625.

[7] R. Beckers, J.L. Deneubourg, S. Goss, 'Trail and U-turns in the Selection of the Shortest Path by the Ant Lasius Niger', *Journal of Theoretical Biology*, **159**,1992, 397–415.

[8] L. Rosati, M. Berioli, G. Reali, 'On Ant Routing Algorithms in Ad-hoc Networks with Critical Connectivity', *Ad-hoc Networks*, **6**,2008, 827–859.
[9] H.J. Byun, M.S. Do, Y. Liu, J. So, R. Taori,'Self-Organization in Future Mobile Communication Networks', White Paper, Wireless World Research Forum (WWRF) Special Interest Group (SIG) 3, Version 1.0, 22 April, 2008.
[10] M. Coates, A.O. Hero, III, R. Nowak, B. Yu, 'Internet Tomography', *IEEE Signal Processing Magazine*, May 2002, 47–65.
[11] M.G. Rabbat, M.A.T. Figueiredo, R.D. Nowak, 'Network Inference from Co-Occurrences', *IEEE Transactions on Information Theory*, **54**(9), 2008, 4053–4068.
[12] Y.C. Peng, *Routing Challenges in Cognitive Radio Networks*, MS Thesis, National Taiwan University, 2008.
[13] E. Tragos, A. Mihovska, E. Mino-Diaz, P. Karamolegkos, 'Access Selection and Mobility Management in a Beyond 3G RAN: The WINNER Approach', *ACM MobiWac*, 2007.
[14] F.S. Chu, K.-C. Chen, 'Radio Resource Allocation for Mobile MIMO-OFDMA', *IEEE Vehicular Technology Conference, Spring*, Singapore, 2008.
[15] C.K. Yu, K.C. Chen, 'Radio Resource Tomography of Cognitive Radio Networks', *IEEE Vehicular Technology Conference, Spring*, 2009.

第10章 可信任认知无线电网络

很明显，认知无线电网络是一种临时组织的同时具有 Ad hoc 网络和异构网络特征的网络，不同的是，认知无线电网络拓扑的更大动态性以及链路的可用性使得静态安全和传统网络功能变得不可行。但是，正如所有的无线网络和因特网应用那样，认知无线电也能被认为是 OSI 七层模型。从最初章节讨论的网络层功能，可知信任是认知无线电网络简化的关键参数之一，因为它的作用是将认知无线电网络中各种各样的异构（无线）网络的性质融合在一起。

在认知无线电网络的运行中，有两个重要的步骤需要信任：联合和路由。这一章我们集中讨论这些主要的问题。当认知无线电最初试图将一个节点连接到现存（认知无线电）网络或者试图形成一个认知无线电网络时，要想执行通常的安全功能是不切实际的，因为

■ 这样的行为可能形成安全漏洞；

■ 在不能确定形成一个认知无线电网络是有效的请求时，为了安全花费大量的计算能力是不明智的；

■ 对于机会性的认知无线电网络链路没有足够的可用时间来进行复杂的握手安全协议。

这就需要一种可信任的机制，鉴定则是信任的一部分，当然还包括其他技术或者非技术的因素。下一个挑战出现在当一个节点试图通过另外一个节点或者另外一个网络的某些部分发送数据流时。典型的 Ad hoc 网络使用公钥基础结构（PKI）方案来达到安全路由和其他目的。但是，由于有限的通信和计算资源下进行核对不现实和增加被攻击的可能性，使得路由包请求下的认知无线电网络节点可能不能执行安全，例如 Ad hoc 网络中公钥基础结构方案。因此，面对这些技术挑战，可信任网络和信任措施的更新达到一个折中的方案是很有用的。其他的安全和服务机制在之后合适的时机也将执行，这也是第9章中讨论（异构）认知无线电网络路由的重要意义。

10.1 认知无线电网络中信任的框架

从社会科学到计算机科学，信任已经被长期广泛地研究。研究认知无线电网络中信任的首要挑战是怎样从数学上定义信任，同时也要注意不信任可能和信任一样重要。根据参考文献［5，6］中的成果，可以观察到认知无线电网络是协

作性的网络，也是异构网络。尽管认知无线电网络可以广义地看作是多跳(Ad hoc)网络，但是它和同构无线 Ad hoc 网络有本质的不同。因此，为了保证认知无线电网络在支持泛在计算时的平稳运行，信任变成了认知无线电网络的基础。在可信任计算和因特网/网页计算、Ad hoc 网络甚至社会科学研究领域，信任被广泛地提及，但是，认知无线电网络的信任和这些应用明显不同。可信任计算处理的是一个集合或者领域内的部件。因特网/网页计算把信任当成某种机制（例如别人的评分）赋予的声誉或者信用。于是必须发展一个数学的框架来为认知无线电网络的信任建模，在认知无线电的设计和运行中量化地应用信任。

因为安全通常需要预先的通信额外开销，所以在认知无线电网络运行中信任远比安全设计重要。以下的例子解释了为什么除了安全还需要信任：

● 一个感知到频谱空洞或者机会，同时动态地访问频谱来传输的认知无线电需要来自原已存在系统（例如主系统）和监管者的信任，即使它没有对主系统造成干扰。

● 一个认知无线电可能想要利用另外一个现存的认知无线电传输数据包，即使另外一个认知无线电不是接收的目标终端。于是这就要求其他认知无线电的“信任”。

● 认知无线电甚至能利用主系统来传递自己的数据包，从而实现数据包交换网络的目的。这需要来自主系统的“信任”，不仅是在网络层，也包括服务提供商（或者网络运营商）。

10.1.1　信任的数学结构

若暂时不考虑运营方，而仅考虑认知无线电网络中的信任。信任必须可测，这样才能使网络基于它运行，正如第 9 章中讨论的每一次路由。直觉上测量信任有如下方法。

定义 10.1：信任是 $[-1, 1]$ 的测度

备注：$\tau(i, j)$ 表示节点 j 处理（接收或者发送）来自节点 i 的信任测量。实际上，是在 $(-\infty, \infty)$ 上测量，并能归一化成 $\tau(i, j) \in (-1, 1)$。归一化的 1 表示全信任（自信）；0 表示关于信任与否没有信息；-1 表示根本没有信任。对于一个认知无线电网络中的决策或者策略，很明显负信任或者零信任得到相同的行为（拒绝数据包）。对定义 10.1 修改如下。

推论 10.1：认知无线电网络中的信任是 $[0, 1]$ 的测度

备注：这正如概率论的测度，使得我们的数学框架可以使用概率推导和统计决策理论。请注意，这个从定义 10.1 产生的变化可能导致对于信任测量在 0 处出现相等的概率质量$\left(例如\int_{-1}^{0} 1_{\tau(i,j)\in[-1,0]}\,d\tau\right)$。

备注：在 0 处的信任测量是不可信任的。任何认知无线电网络中的节点一旦

被认为是只有0信任，它将被任何认知无线电网络的行为所拒绝。也即任何连接到该节点上的链路都要从认知无线电网络中移除。测度为0或者为负没有任何意义。

引理10.1：认知无线电网络中的信任通常是不可逆的，也即

$$\tau(i,j)\neq\tau(j,i) \tag{10-1}$$

备注：对于任何结果采用信任的概念通常都是对的。Alice（爱丽丝）信任Bob（鲍勃）的程度和Bob信任Alice的程度是不等的。

定义10.2：（度量空间[7]），每个标准空间可以认为是一个度量空间，$d(x,y)$代表x和y之间的距离，有如下特性：

（i） $0\leqslant d(x,y)<\infty\ \forall x,y$

（ii） $d(x,y)=0$，当且仅当$x=y$

（iii） $d(x,y)=d(y,x)\ \forall x,y$

（iv） $d(x,z)\leqslant d(x,y)+d(y,z)\ \forall x,y,z$

引理10.2：认知无线电网络中的信任不是一个度量。

证明：为了构造一个度量空间，$\tau(i,j)\geqslant 0$，这可以通过引进偏倚来求解。但是，$\tau(i,j)+\tau(j,k)\leqslant\tau(i,k)$违反了度量空间的要求。这个不等式的含义是通过一个中间节点的信任不比直接从起点的信任高。不仅如此，信任是不可逆的，也即，$\tau(i,j)\neq\tau(j,i)$。这可以通过节点i是一个移动的认知无线电、节点j是蜂窝网络的基站的情况来解释。

备注：这个推论在大多数信任建模的研究领域文献中都被忽略了。但是，在这里它很关键。因为许多信任测量系统都是在信任测量是度量的假设上建立的，在进一步的发展中，这样的基本假设可能产生风险。

备注：但是，若定义“不信任”来取代信任（例如，D作为不信任的测量），三者之间的不等式为

$$D(i,j)+D(j,k)\geqslant D(i,k) \tag{10-2}$$

正如很多作者所做的一样，必须指出，在网络研究中，给不信任建模的重要性可能并不比给信任建模次要（不管是因特网还是认知无线电网络）。

定义10.3：认知无线电网络中的信任来自声誉（其他节点的信任测量）和协同（目的节点或者其他节点的观测行为）。任何0信任意味着不信任。

备注：声誉是一个从电子商务信任中借来的术语。一个CR节点可以通过在运行规则允许的情况下执行更多的行动来增加声誉。声誉和协同都遵循定义10.1。但是声誉或者协同的任何一部分若等于0，则意味着0/没有信任。定义10.3和研究文献中通常的加性定义不同，建议“相乘”（或者半群）更合适。

命题10.1：（信任路径理论）认知无线电网络中的信任是路由路径的函数。总信任是每个部分信任的乘积，即

$$\tau(n_0, n_1, \cdots, n_L) = \prod_{l=0}^{L-1} \tau(n_l, n_{l+1}) \tag{10-3}$$

$$\tau(n_0, n_1, \cdots, n_L) \neq \tau(n_0, n_L) \tag{10-4}$$

备注：对于 3 个节点（2 个部分）的数据包传输

$$\tau(i, j, k) = \tau(i, j)\tau(j, k) \neq \tau(i, k)$$

备注：在文献中，许多研究者提到信任取决于以前的历史，这和命题 10.1 是相符的。不过，命题 10.1 指出了信任不仅仅是历史，应该是历史路径的级数。

命题 10.2：（信任处理理论）在 CRN 中，信任处理（例如数据包/数据流传输）不会增加信任本身。

备注：这是定义 10.2 和命题 10.2 的直接结果。在认知无线电网络中，很容易理解：对于任何从源节点出发的数据包，在经历多个环节的传输后就没有信任了。

定义 10.4：（半群[7]）假定 X 是 Banach（巴拿赫）空间，且假设对于每个 $t \in [0, \infty)$ 都有对应的运算符 $Q(t) \in B(X)$，并且

（ⅰ）$Q(0) = I$

（ⅱ）$Q(s+t) = Q(s)Q(t)\ \forall s,\ t \geqslant 0$

命题 10.3：同构 Ad hoc 网络中的信任减弱为半群，如参考文献［8］所述。

备注：请注意，这里要求的条件是同构 Ad hoc 网络，因为 Ad hoc 网络意味着信任的可逆（或交换）成立，如式（10-1）所述。但是，认知无线电网络通常既不是同构的也不是 Ad hoc（认知无线电网络在某些部分可能有基础设施）。从这个数学特性能容易地知道认知无线电网络中的信任和 Ad hoc 网络中的信任的区别。

要在认知无线电网络中识别一个信任的数学测度可能是有争议的。但是，大家广泛同意的是信任测度有和概率测度一样的性质，正如推论 10.1 指出的。让我们从人工智能的一个分支——数学推理开始。

引理 10.3：（Cox 公理）我们想要测量和以下条件一致的任何“确定性”：

确定性的程度可以排序；

存在一个函数将声明的确定性映射到它的反/补；

推论 $R(A, B)$ 的程度通过函数 g 与条件推论 $R(A|B)$ 及 $R(B)$ 相关

$$R(A \wedge B) = g(R(A|B), R(B))$$

因此，这个推论系统必然与概率测度等价。

备注：信任模型在认知无线电网络中确实是可用的。因此，之后我们可以像概率一样测量信任。

10.1.2 信任模型

认知无线电中信任模型的主要目的是提供一套数学框架，以在异构网络中实现感知、测量、分析和学习拓扑的变化以及邻居的行为。在这样的网络里，信任

模型在不同系统的实体间起了很重要的作用，它提供了节点间建立信任连接的机制。在建立信任关系后，实体之间的协作成为可能，因此它们可以互相通信，一起工作，例如数据交通中转、协作者选择和信任路由。本章里，我们将为认知无线电建立信任模型，这为各节点提供以下两个元素：

■ 信任连接：一个节点的初始决策是接收还是拒绝一个邻居认知无线电节点的信任连接请求。这个初始决策的核心概念是在最小风险下实现联网。

■ 学习算法：认知无线电里的每一个节点都应该保持其他节点的记录，采用一种适当的学习算法来适应数据包转发的概率/信任度量的长期趋势，以对保持被信任路由或收回进行决策。

我们在图 10-1 中描述提出的信任模型的基本元素，图中描述了两个功能之间的相互影响，并且提供信任模型的整个场景。在无线网络中，信道条件总是在改变，因此信任在认知无线电网络里也是动态的。我们需要发展出一套适宜的学习算法，来适应在各种情况下与邻居间协作的趋势。认知无线电网络里的每一个节点可以根据现存的环境决定其关于邻居的信任策略。该算法必须能够捕捉恶劣行为，对企图损害网络流量的节点进行惩罚。另一方面，这个机制需要给这些节点从被隔离的状态下恢复的机会。

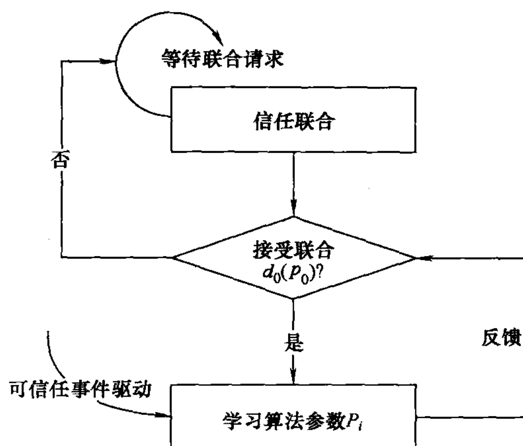


图 10-1 CRN 中的信任模型流程

10.2 可信任联合和路由

CRN 其中一项最困难的任务是一个 CR（发送节点）向协作主系统的一个节点或另一个 CR 节点（接收节点）发起联合请求，执行认知无线电网络功能。和第 4 章所介绍的协作网络相似，CRN 中的接收节点可以执行以下其中一项：

1）拒绝联合；2）放大转发（AF）模式；3）压缩转发（CF）模式。AF 和 CF 的不同之处在于 AF 只是执行物理层的功能，因而我们不需要担心受攻击，而 CF 实际上在潜在的安全威胁下将数据包解码传给上层。换言之，AF 错误判决的代价只是浪费了通信带宽而已（也可能是电池能量），而 CF 错误判决的代价可能是损害整个网络，因而提出在用 CF 方式联合之前使用诸如 PKI 这样的安全校验。

10.2.1 可信任联合

我们在图 10-2 中描述了 CRN 的联合，并采用 Neyman-Pearson 准则（由于判决没有使用任何先验概率或代价函数）。

命题 10.4：基于信任度量，一个 CRN 中的联合只可能有两种判决，接受联合或者拒绝联合。设 X 表示信任度量值，其分布 $F_\theta(x)$ 表示从 CR-MS 向 PS-MS发出的联合请求中的信息。设 $\Theta=\Theta_0\cup\Theta_1$ 是对信任空间的一个不连续覆盖，而 H_i表示假设参数 θ 属于信任空间 $\theta\in\Theta_i$。现在，判决问题是通过考虑如 6.2 节中的 CRN 的架构来区分如下两个假设：



图 10-2　信任判决中应用 Neyman-Pearson 准则

1）H_0：$\theta\in\Theta_0$：这表示 CR-MS 将不值得信任。CR-MS 的概率密度函数为 $f_{\tau|0}=f_\tau(x\mid 0)$，其中 x 表示 CR-MS 的信任度量。

2）H_1：$\theta\in\Theta_1$：这表示 CR-MS 值得信任。CR-MS 的概率密度函数为$f_{\tau|1}=f_\tau(x\mid 1)$，其中 x 表示 CR-MS 的信任度量。

如果从 CR-MS 得到的信任度量 x 对应的概率密度函数在 $\theta\in\Theta_1$的信任空间下比在信任空间 $\theta\in\Theta_0$下大时，PS-MS 判决信任 CR-MS。

备注：对于 AF，一旦接受联合，数据包将被中继。对于 CF，一旦接受联合，发送节点必须通过安全检测，然后接收节点压缩并中继数据包。

命题 10.5：当 CRN（主系统或次要系统）中的一个节点从一个新节点接收到联合的请求（即表示要加入该 CRN），它会做出如下统计判决：基于该点对应的信任度量 τ，做出判决 $\delta(x)=a_i$，其中 a_1表示“接受”联合，而 a_0表示“拒绝”联合。

如果从 CR-MS 得到的信任度量 x 对应的概率密度函数在 $\theta\in\Theta_1$的信任空间下比在信任空间 $\theta\in\Theta_0$下大时，PS-MS 判决信任 CR-MS。因此根据这两个假设，我们可以对信任度量制定如下判决准则：

$$\delta(x)=\begin{cases}1,\ f_{\tau|1}(x|1)>\gamma f_{\tau|0}(x|0)\\ k,\ f_{\tau|1}(x|1)=\gamma f_{\tau|0}(x|0),\ \gamma\geqslant 0,\ 0\leqslant k\leqslant 1\\ 0,\ f_{\tau|1}(x|1)<\gamma f_{\tau|0}(x|0)\end{cases}\tag{10-5}$$

给定虚警概率使检测概率最大的判决是根据 Neyman-Perarson 准则做的似然比检测。为了在给定一个 $P_F \leqslant \alpha$，要使 P_D最大，PS-MS 判决信任 CR-MS，如果

$$l(x) = \frac{f_{\tau|1}(x|1)}{f_{\tau|0}(x|0)} > \gamma \tag{10-6}$$

同理，PS-MS 判决不信任 CR-MS，如果

$$l(x) = \frac{f_{\tau|1}(x|1)}{f_{\tau|0}(x|0)} < \gamma \tag{10-7}$$

似然比函数

$$l(x) = \frac{f_{\tau|1}(x|1)}{f_{\tau|0}(x|0)}$$

指示在每次信任度量下选择 H_1 的概率对比选择 H_0 的概率。我们可以将判决换一种形式为

$$\delta(x) = \begin{cases} 1, & l(x) > \gamma \\ k, & l(x) = \gamma,\ \gamma \geqslant 0,\ 0 \leqslant k \leqslant 1 \\ 0, & l(x) < \gamma \end{cases} \tag{10-8}$$

备注：我们可以建立一个二元统计判决问题的模型来描述成功联合的过程。这可以应用在更为实际的 CRN 动态频谱接入的研究中。请同时注意，随机判决准则在此仍有可能使用。然而，为了完整定义二元判决问题，我们仍需要引入代价函数和先验概率（信任度量）分布的定义。由于很难在实际中建立可信任联合之前设定代价函数或先验概率，因此我们采用 Neyman-Pearson 准则来解决这个问题。

定义 10.5：根据之前的命题，我们可以根据对信任联合的理解，定义虚警概率和检测概率。我们首先定义给定以 CR-MS 不信任 PS-MS 作为虚警概率，PS-MS 信任 CR-MS 的概率为

$$P_F = \int_{x \in Z_1} f_\tau(x|0)\,\mathrm{d}x \tag{10-9}$$

接着，我们定义给定 CR-MS 信任 PS-MS 作为检测概率，PS-MS 信任 CR-MS 的概率为

$$P_D = \int_{x \in Z_1} f_\tau(x|1)\,\mathrm{d}x \tag{10-10}$$

备注：我们可以容易看出信任度量的重要性。信任度量由正确的信任和观测的偏差（例如一般通信理论中的观测噪声）组成。这样的观测偏差趋向于负的分布，这是因为从命题 10.2（信任处理理论）中可以知道更多的观测值不能增加信任值。然而，对于恶意节点，这样的结论不一定成立，而且观测偏差在分布中可以是双边的。

我们在这个判决问题中使用 Neyman-Pearson 准则是为了在给定 CR-MS 不值得信任的约束条件下，使得 PS-MS 信任 CR-MS 的风险最小化。在此，我们不对

信任度量的概率密度函数作任何假设。它可能是离散的或连续的，而我们只需要针对提出的信任模型给出一个通用的信任判决准则。

命题 10.6：当信任度的先验信息未知，同时代价函数不能准确地定义时，可以根据 Neyman-Pearson 准则作出判决 $\delta(x)=a_i$。也就是说，给定 $P_F\leqslant\alpha(0\leqslant\alpha\leqslant1)$，优化 P_D。

备注：给定虚警概率使检测概率最大化的方法是似然比检测。对于 Neyman-Pearson 检测，P_F和 P_D的值完全决定了检测的性能。信任判决问题转化为

$$\text{最大化 } P_D\text{，当 } P_F=\alpha \tag{10-11}$$

$$\text{判决 } \delta(x)=\begin{cases}1, & l(x)>\gamma\\ k, & l(x)=\gamma,\ 0\leqslant k\leqslant1\\ 0, & l(x)<\gamma\end{cases} \tag{10-12}$$

式中，判决门限 γ 由具体的系统约束来决定：

$$P_F=\int_{\{x:l(x)>\gamma\}}f_{T|0}(x|0)\,\mathrm{d}x=\alpha \tag{10-13}$$

如果信任是基于个人观测，很可能不同的观察者观测到不同的情形；因而 CRN 中各个节点根据它对恶意行为的容忍水平将采纳不同的显著性水平。增大 γ 可以减小检测对于干扰的敏感度，而我们接受更高的检测概率作为较小虚警概率的回报。

我们现在给出 CRN 中的节点如何应用以上所讨论的 Neyman-Pearson 准则作判决的一个例子。设信任度量服从正态分布，以便在此实际地解决信任问题。考虑图 10-3 的情况，其中 CR-MS-A 独立地以概率 p 中继/转发数据包，并以概率 $1-p$ 丢弃/忽略从 CR-MS-B 接收到的数据包。根据这个信息，PS-MS 必须接受或拒绝从 CR-MS-A 发出的联合请求，使得在系统约束 P_F下最大化 P_D。

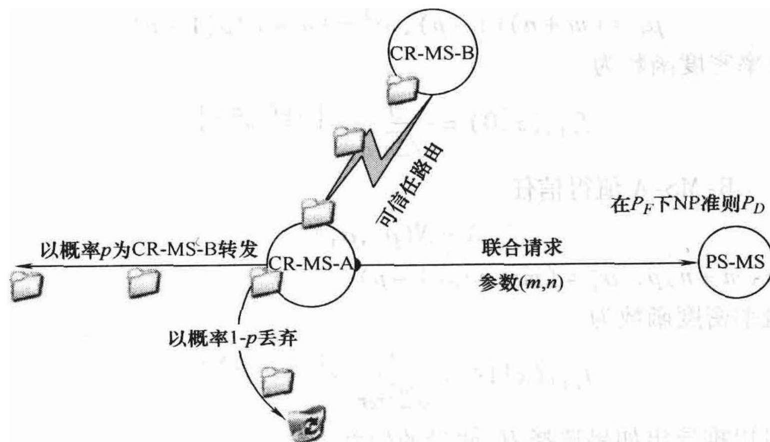


图 10-3 基于 Neyman-Pearson 准则（应用正态分布）实行联合

设 X 为 CR-MS-A 转发的数据包个数，并且我们假设每个数据包独立地转发或被丢弃。这个假设是合理的，因为在网络中传输的数据量足够大。接着，在这个场景中，X 服从参数为（$m+n$，p）的二项分布，其中 p 表示成功的概率。由于 $m+n$ 足够大，并且根据 De Moivre-Laplace 定理，我们可以知道对于任意的 a 和 b，$a<b$，有

$$\lim_{n\to\infty} P\left(a < \frac{X-(m+n)p}{\sqrt{(m+n)p(1-p)}} < b\right) = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-t^2/2}\mathrm{d}t \tag{10-14}$$

式中，期望为

$$\mathrm{E}(X) = (m+n)p \tag{10-15}$$

而成功概率的方差为

$$\sigma_X = \sqrt{(m+n)p(1-p)} \tag{10-16}$$

由于数据包被成功发送的概率在下一级可能不能有显式解，我们可以根据经验等级近似得到这个概率：

$$Pr(\text{协作}) \approx \frac{\text{数据包转发量}}{\text{数据包接收量}} \tag{10-17}$$

由这一近似解，CR-MS-A 成功转发数据包的概率为 $\frac{m}{m+n}$，我们将它定义为可信任协作概率。我们可以在这个问题中用正态分布来建立 CR-MS-A 转发数据包这一操作的模型。现在，问题转化为以下假设问题：

■ H_0：CR-MS-A 不值得信任。

$$X \sim N(\mu_0,\ \sigma_0^2)$$

式中

$$\mu_0 = (m+n)(1-p),\ \sigma_0^2 = (m+n)p(1-p)$$

x 的概率密度函数为

$$f_{\tau|0}(x|0) = \frac{1}{\sqrt{2\pi}\sigma_0}e^{\left\{-\frac{1}{2}\left(\frac{x-\mu_0}{\sigma_0}\right)^2\right\}} \tag{10-18}$$

■ H_1：CR-MS-A 值得信任。

$$X \sim N(\mu_1,\sigma_1^2)$$

式中，$\mu_1=(m+n)p$，$\sigma_1^2=(m+n)p(1-p)$

X 的概率密度函数为

$$f_{\tau|1}(x|1) = \frac{1}{\sqrt{2\pi}\sigma_1}e^{\left\{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right\}} \tag{10-19}$$

我们可以推导出如果选择 H_1 假设的似然比为

$$l(x)=\frac{f_{\tau|1}(x|1)}{f_{\tau|0}(x|0)}=\frac{\frac{1}{\sqrt{2\pi}\sigma_1}e\left\{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right\}}{\frac{1}{\sqrt{2\pi}\sigma_0}e\left\{-\frac{1}{2}\left(\frac{x-\mu_0}{\sigma_0}\right)^2\right\}}>\gamma \tag{10-20}$$

最后，它等价于

$$x>\frac{\sigma_1^2}{\mu_1-\mu_0}\ln\gamma+\frac{\mu_1+\mu_0}{2}=\gamma \tag{10-21}$$

我们可以由虚警约束决定判决门限：

$$P_F=Pr\{x>\gamma'|H_0\}\leqslant\alpha \tag{10-22}$$

然后，PS-MS 可以做出使检测概率最大的判决：

$$\begin{aligned}P_D&=Pr\{\text{在给定 } H_1 \text{ 的条件下判决 } H_1\}\\&=Pr\{x>\gamma'|H_1\}\\&=\int_{\gamma'}^{\infty}f_{\tau|1}(x|1)\,dx\\&=Q\left(\frac{\gamma'-\mu_1}{\sqrt{\sigma_1^2}}\right)\end{aligned} \tag{10-23}$$

在系统约束下，由虚警概率也可以推导出判决门限：

$$\begin{aligned}P_F&=Pr\{\text{在给定 } H_0 \text{ 的条件下判决 } H_1\}\\&=Pr\{x>\gamma'|H_0\}\\&=\int_{\gamma'}^{\infty}f_{\tau|0}(x|0)\,dx\\&=Q\left(\frac{\gamma'-\mu_0}{\sqrt{\sigma_0^2}}\right)\end{aligned} \tag{10-24}$$

式中，$Q(x)$ 是零均值单位方差的高斯随机变量的右尾概率：

$$Q(x)=\int_x^{\infty}\frac{1}{\sqrt{2\pi}}e^{-t^2/2}dt \tag{10-25}$$

应用约束条件 $P_F=\alpha$，我们可以推导出判决门限的真值：

$$\gamma'=\sigma_0Q^{-1}(\alpha)+\mu_0 \tag{10-26}$$

而且，我们可以得到检测概率为

$$\begin{aligned}P_D&=Q\left(\frac{\sigma_0Q^{-1}(\alpha)-(\mu_1-\mu_0)}{\sqrt{\sigma_1^2}}\right)\\&=Q\left(Q^{-1}(\alpha)-\frac{(\mu_1-\mu_0)}{\sqrt{\sigma_1^2}}\right)\end{aligned} \tag{10-27}$$

如果我们定义系数 c^2：

$$c^2=\frac{(\mu_1-\mu_0)^2}{\sigma_1^2}=\frac{(m+n)(2p)^2}{p(1-p)} \tag{10-28}$$

作为等同于在大多数检测问题中用来近似检测性能的评估的偏转系数，我们可以从这一系统约束条件中推导出可信任协作的最大概率：

$$P_D=Q(Q^{-1}(\alpha)-\sqrt{c^2}) \tag{10-29}$$

给定虚警概率作为约束条件，由此推出判决门限的真值，最优信任判决可以表示为

- 如果从邻近可信任协作中最大化 P_D，就判为可信任的。
- 如果从邻近可信任协作中不能最大化 P_D，就判为不可信任的。

10.2.2 可信任路由

一旦一个节点在联合后被 CRN 接受，正如典型多跳网络中，CRN 将更新它的网络拓扑和路由表。如果我们把 CRN 看成一个同构放大转发网络，路由将和其他多跳 Ad hoc 网络相同。然而，CRN 显然不是同构的。安全性作为 CRN 中特别严重的风险，开发出一套可信任的物理层功能是十分重要的，特别是要考虑拓扑和路由。基于已经提出的 CRN 信任问题的数学框架，我们可以得到 CRN 在网络层的基本操作，在 CRN 中网络拓扑建立和路由。

在一个同构网络中典型的路由算法考虑的是反映可用带宽/容量和传输代价的距离度量。对于 CRN 的路由，我们必须考虑在异构网络环境中的信任问题（例如，PS 是蜂窝网络，而 CR 是 Ad hoc 网络 WiFi 基站）。

命题 10.7：（可信任路由）在 CRN 中的节点 i 和节点 j 的路由测量是由一个状态为（$\tau(i,\ j)$，$d(i,\ j)$）的状态机来定义的，它表示了信任度量和距离度量。

为了有效利用诸如 Dijkstra 算法等路由算法，我们可以假设任意两点间存在可逆的信任关系，简单的定义一个新的信任距离 $D(i,\ j)=d(i,\ j)/\tau(i,\ j)$，来重复算法找出路由。在没有信任度（$\tau(i,\ j)=0$）的情况下，链路由于无穷距离而被消除。图 10-4 描述了存在互逆信任关系，应用新的信任距离度量的 Bellman-Ford 和 Dijkstra 路由算法。

在两个节点之间的信任关系不互逆，也就是说两个方向的距离不相同时的路由仍旧是一个有待解决的问题，这是第 9 章中讨论过的 CRN 中单向链路的非对称性。

定义 10.6：一个 CRN 节点（它可以是连接 PS 的接入点），决定信任度并对来自另一个 CRN 节点的数据包采取行动，这可形成一个马尔可夫判决过程（给定判决规则和状态空间）[5]。

命题 10.8：（马尔可夫信任过程）对于数据包从节点 i 传输到节点 j，接收

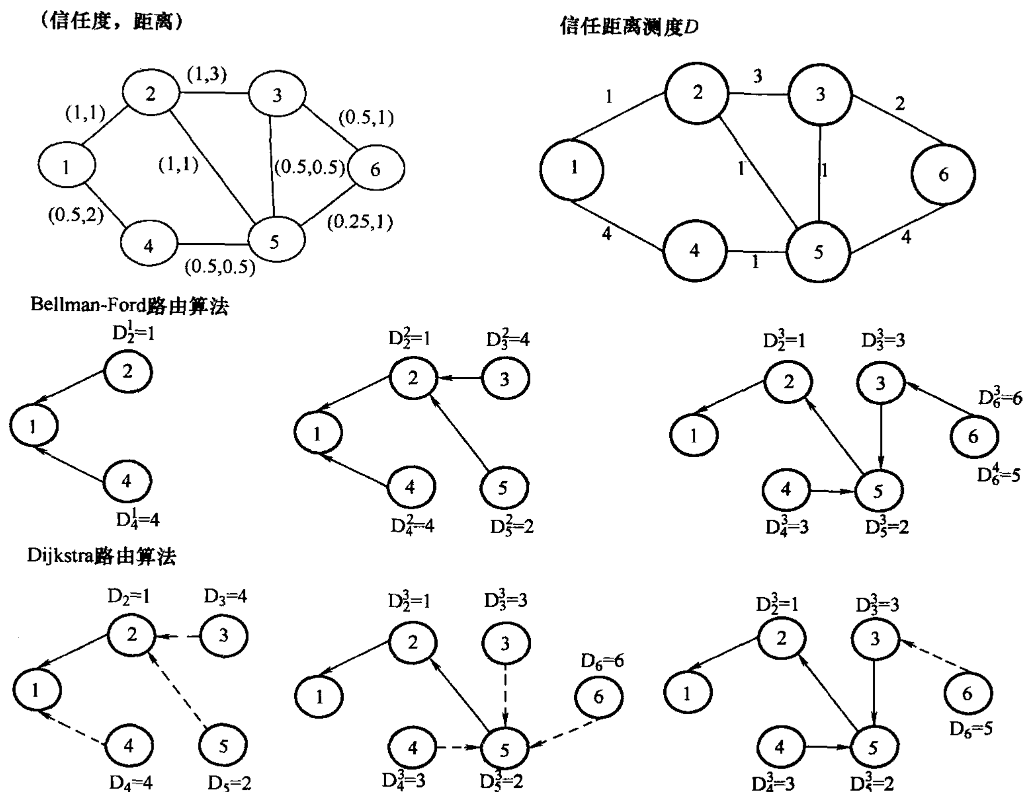


图 10-4　可信任路由举例

节点（在此是节点 j）可以基于某种判决规则构成一个二元假设检验（H_1 表示信任，而 H_0 表示不信任）。命题 10.8 中的可信任路由变为一个马尔可夫过程。

备注：可信任路由因此成为一种马尔可夫决策过程。事实上，在 CRN 中随机判决信任度是可能且是有意义的。例如，由于不同的原因，诸如，用户/节点漫游、对抗攻击的鲁棒性或只是由于没有足够的信用额度，$\tau(i,\ j)=\tau_r$，$0<\tau_r<1$ 可能表示一个信任状态的贴现因子。状态的个数可以是有限的或无限的。

10.3　信任学习

认知无线电网络的信任的真正挑战不仅是构建信任的措施去采取适当的行动，但其后更是为了每个网络节点进行可能的通信/网络功能，如数据包的请求中继，而更新信任的“分布函数”。主系统中支持认知无线电网络的每个节点或者每个认知无线电应该能够更新和维持相邻节点的（为之后章节所描述的）应用的一个信任列表。虽然在数学上衡量认知无线电网络的信任可能是一个有争议

的问题，因此建议信任的衡量是一个概率的测度。基于这样的衡量，我们可以在不同的应用情景以信任衡量为基础进行判决。在本节中，我们着重论述如何更新信任衡量的分布函数。

10.3.1 改进的贝叶斯学习

假设信任的协作概率分布是$f_p(p \mid \alpha_t, \beta_t)$，$0 \leqslant p \leqslant 1$，其中下标$t$代表离散时间索引，$t=0$是初始分布，$f_p(p \mid \alpha_t, \beta_t)$能被$f_p(p \mid \alpha_{t-1}, \beta_{t-1})$基于当前的信任证据、观察和任何未来的信息而递归更新。为了解决这个挑战，我们可为认知无线电网络节点而采用一种学习策略去适应在每个时间点数据包从相邻点传递的概率分布。

引理10.4：(改进的贝叶斯学习策略）假设贝塔密度f_p（$p \mid \alpha_{t-1}, \beta_{t-1}$）在我们的模型中代表在时间$t$的信任协作的概率分布。学习策略包括三部分：概率预测、衰减校正和测量修改：

$$\hat{p}_t = \int p(x_t \mid P=p) f(p \mid \alpha_t, \beta_t)\,\mathrm{d}p \tag{10-30}$$

$$f_{t-1}(\hat{p}) = f(p \mid k_1\alpha_{t-1}, k_2\beta_{t-1}) \tag{10-31}$$

$$f(p \mid \alpha_t, \beta_t) = cf((m_t, n_t) \mid p) f_{t-1}(\hat{p}) \tag{10-32}$$

式中，t是离散时间索引，k_1和k_2是衰减因子，(m_t, n_t)是新信任证据，c是改进因子代表恒定的整数，$\hat{p}_t$是信任协作的预测概率。

备注：预测是用来预测在下一个阶段的信任协作的概率测量，它首要是为决策准则而设计。引理10.4包括了更新概率分布和更新规则。概率密度函数$f(p \mid \alpha_t, \beta_t)$为预测函数合并了新信任证据和一个先验概率密度函数。它包括了在$T=t-1$之前的所有信息，其中包括了初始值的所有信任证据在接收到新数值时进行衰减：

$$\begin{aligned}\alpha_t &= k_1\alpha_{t-1} + m_t = k_1^t m_0 + k_1^{t-1} m_1 + k_1^{t-2} m_2 + \cdots + k_1 m_{t-1} + m_t \\ \beta_t &= k_2\beta_{t-1} + n_t = k_2^t n_0 + k_2^{t-1} n_1 + k_2^{t-2} n_2 + \cdots + k_2 n_{t-1} + n_t\end{aligned} \tag{10-33}$$

然后，我们通过式（10-31）和在式（10-32）引入一个新信任测量后进行在时间$T=t$之前的过去信息的衰减校正。我们用衰减校正“忘记”过去信任的证据来解释有效期的限制，因为该信任模型要逐步忽视最老的记录以捕捉最新的记录。我们使用两个常数K_1和K_2代表随着时间推移的衰减因子，通常K_1是小于K_2的，以捕捉不良行为的节点，如欺骗或不良意图的丢包。

当我们在时间$T=t$收到新的信任证据时，为了支持在认知无线电网络的动态和快速操作，给最新记录更高的权重是适当的。因此，为了信任认知无线电网络，通过这种学习机制我们可以维持参数的更新。

备注：在推导预测概率$\hat{p}_t$之前，我们需要描述校正公式的细节。在时间$T=t$

结束时，我们有新的信任证据（m_t，n_t）和先验概率分布，我们能计算出概率密度函数如下：

$$
\begin{aligned}
f(p|\alpha_t,\beta_t) &= \frac{f_\tau((m_t,n_t)|p)f_{t-1}(\hat{p})}{\int f_\tau((m_t,n_t)|p)f_{t-1}(\hat{p})\mathrm{d}p} \\
&= \frac{p((m_t,n_t)|P=p)f(p|k_1\alpha_{t-1},k_2\beta_{t-1})}{\int p((m_t,n_t)|P=p)f(p|k_1\alpha_{t-1},k_2\beta_{t-1})\mathrm{d}p} \\
&= \begin{cases} \dfrac{\Gamma(\alpha_t+\beta_t)}{\Gamma(\alpha_t)\Gamma(\beta_t)}p^{\alpha_t-1}(1-p)^{\beta_t-1}, & 0\leqslant p\leqslant 1 \\ 0, & 其他 \end{cases}
\end{aligned} \tag{10-34}
$$

我们提供了学习机制的技术流程图，如图 10-5 所示。这对我们的信任模型提供了两个重要的概念。首先是时间尺度的测量校正。过去的信任证据必须衰减以支持认知无线电网络中的即时互动，并提供一个从坏的通道状态的恢复方法。第二是预测。信任的判决是在每一个阶段，主要是基于这一预测值；是否持有或收回它。在我们的方法中给出的预测功能不算是学习算法的唯一解决办法，它取决于选择的系统模型。在收到新的信任证据（m_t，n_t）后，认知无线电网络的每个节点根据过去的纪录和最新的信任的证据可以预测下一阶段信任协作的可能性。在这种学习模式中，我们所尝试的是在给定概率分布 p、$f_p(p)$ 和新信任证据（m_t，n_t）推断概率 $\hat{p}_t$。

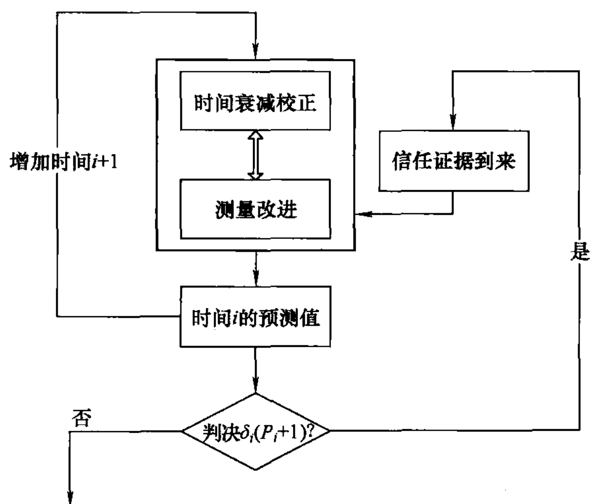


图 10-5　学习机制的流程图

下一个传输成功的概率依赖于信任协作的预测概率，$P=\hat{p}_t$，然后我们能计算出概率：

$$P(x_t|P=\hat{p}_t)=\begin{cases}\hat{p}_t, & x_t=1\\ 1-\hat{p}_t, & x_t=0\\ 0, & \text{其他}\end{cases} \tag{10-35}$$

当我们在时间 t 接收到新的信任证据时，为了支持认知无线电网络的动态和快速操作，给予最新的证据、更高的权重是合适的。最终，我们能直接计算预测概率 $P(x_t=1)$：

$$\begin{aligned}P(x_t=1) &= \int P(x_t=1\mid P=\hat{p}_t(m_t,n_t))f(p\mid \alpha_{t-1},\beta_{t-1})\mathrm{d}p\\ &= \int P(x_{t+1}=1\mid P=p)f(p\mid \alpha_t,\beta_t)\mathrm{d}p\\ &= \int pf(p\mid \alpha_t,\beta_t)\mathrm{d}p\\ &= \frac{\Gamma((k_1\alpha_{t-1}+m_t)+(k_2\beta_{t-1}+n_t))}{\Gamma(k_1\alpha_{t-1}+m_t)\Gamma(k_2\beta_{t-1}+n_t)}\int_0^1 p^{k_1\alpha_{t-1}+m_t}(1-p)^{k_2\beta_{t-1}+n_t-1}\mathrm{d}p\\ &= \frac{k_1\alpha_{t-1}+m_t}{(k_1\alpha_{t-1}+m_t)+(k_2\beta_{t-1}+n_t)}\end{aligned} \tag{10-36}$$

我们重写公式从另外一个形式去表示，它是给定了（m_t，n_t）的最大估计 $P=p$ 的权重均值和先验信息的均值：

$$\begin{aligned}P(x_{t+1}=1)=&\left(\frac{m_t+n_t}{k_1\alpha_{t-1}+k_2\beta_{t-1}+m_t+n_t}\right)\frac{m_t}{m_t+n_t}+\\ &\left(\frac{k_1\alpha_{t-1}+k_2\beta_{t-1}}{k_1\alpha_{t-1}+k_2\beta_{t-1}+m_t+n_t}\right)\frac{k_1\alpha_{t-1}}{k_1\alpha_{t-1}+k_2\beta_{t-1}}\end{aligned} \tag{10-37}$$

当我们获得每个阶段的预测时，就能够根据决定准则判决：

$$\delta_t(m_t,n_t)=\begin{cases}0, \ \hat{p}_{t+1}\leqslant\gamma_i \ \text{表示“拒绝”}\\ 1, \ \hat{p}_{t+1}>\gamma_i \ \text{表示“接受”}\end{cases} \tag{10-38}$$

如果信任协作的概率测量大于阈值 γ_i，这意味下一个阶段数据包能被成功递送，而不是被丢弃具有更大的概率，反之亦然。

10.3.2 认知无线电网络的学习实验

我们可以描述一些场景来说明认知无线电网络学习算法的特性，当为认知无线电网络建立了信任模型，我们可以进一步总结一些一般性的规则。由于认知无线电网络是一个高度动态的异构网络，节点可以动态地、即时地离开或加入网络，认知无线电网络的拓扑结构也可以非常频繁地改变。学习算法应及时地跟踪信道变化和用户行为去学习数据包递送行为中的更新（有利或不利的）变化。

10.3.2.1　节点断开与初值影响

节点断开是认知无线电网络中经常遇到的，学习算法应能尽快处理这个极端的情况以免恶化网络的信任拓扑。正如图 10-6 所示，我们在时间 $t=0$、$k_1=0.2$、$k_2=0.5$ 以信任协作的概率 $p=0.8$ 接受来自邻近节点的信任协作。在时间 $t=21$，邻近节点从网络断开，学习策略立即得到预测的概率，因为信任协作的概率快速下降。在时间 $t=22$，我们收回信任协作，在时间 $t=27$，判决节点已经离开了网络。因为认知无线电网络中判决是动态的，它们也许瞬间离开或加入网络，我们在确定从网络中断开之前只能再观察多一些时间。在图 10-7 中，除了协作请求信息不同外，我们显示了同样的情况。在图 10-6 中数据包的总和是 100，在图 10-7 中数据包的总和是 1300。我们标记初始值影响了接下来在时间 $t=3$ 和 $t=4$ 的信任决策。即使我们立即学习了节点的行为，我们也应该假设初始值去避免模型中的这种情况。为了完全得到最新的节点信任证据，之后的仿真应该采用一个初始值 $m_0+n_0=100$。现在，我们从这个例子提出拇指第一规则：

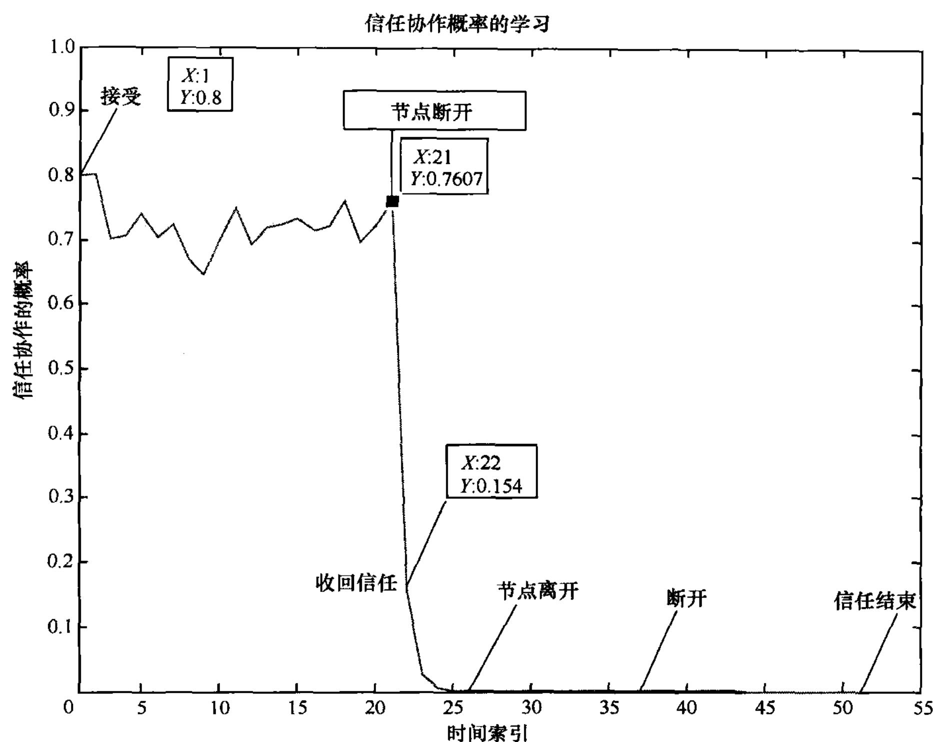


图 10-6　当 $m_0+n_0=100$，节点从网络中断开

图 10-7 当 $m_0 + n_0 = 1300$，节点从网络中断开

命题 10.9：（在不同业务模式的拇指法则）学习算法引理 10.4 在繁忙传输密度下也可以快速学习新的信任证据。即使初始值很大或环境经常变化，如果传输繁忙和衰减因子足够小，我们仍然可以快速学习以获得最新的信任证据。

10.3.2.2 节点快速离开和加入网络

邻近节点发送一个协作请求包括过去在时间 $t=1$，$k_1=0.2$，$k_2=0.5$ 下以信任协作概率 $p=0.8$ 的 100 个的记录。节点在时间 $t=21$ 离开，在时间 $t=32$ 返回。在时间 $t=22$ 时，由于大量的数据包被丢掉，我们立即收回信任的路由，并宣布该节点已离开网络。然而，当节点返回时，我们看到了信任协作的可能性相当大，我们不立即重新建立信任路由。我们观察更多的时间来建立信任路由。这是用来惩罚网络性能下降，并确保节点不会返回和再次离开。我们注意到节点也许在时间 $t=35$ 时返回，所以在时间 $t=36$ 重建信任路由。除了在每个时段传输密度是不同的，图 10-8 所示情形如图 10-6 和图 10-7 相同。图 10-8 的传输密度是 100，图 10-9 的传输密度是 30，它分别代表了繁忙和轻闲的传输网络。从图

10-8 中，我们观察到初始值显著地影响了学习算法，因为初始值之和为 100，对于传输来说它是比较大的。因此，在时间 $t=6$ 和 $t=7$，信任判决被收回，并以连续的方式重新建立。我们现在通过观察初始值问题的影响，准备提出其他拇指规则。

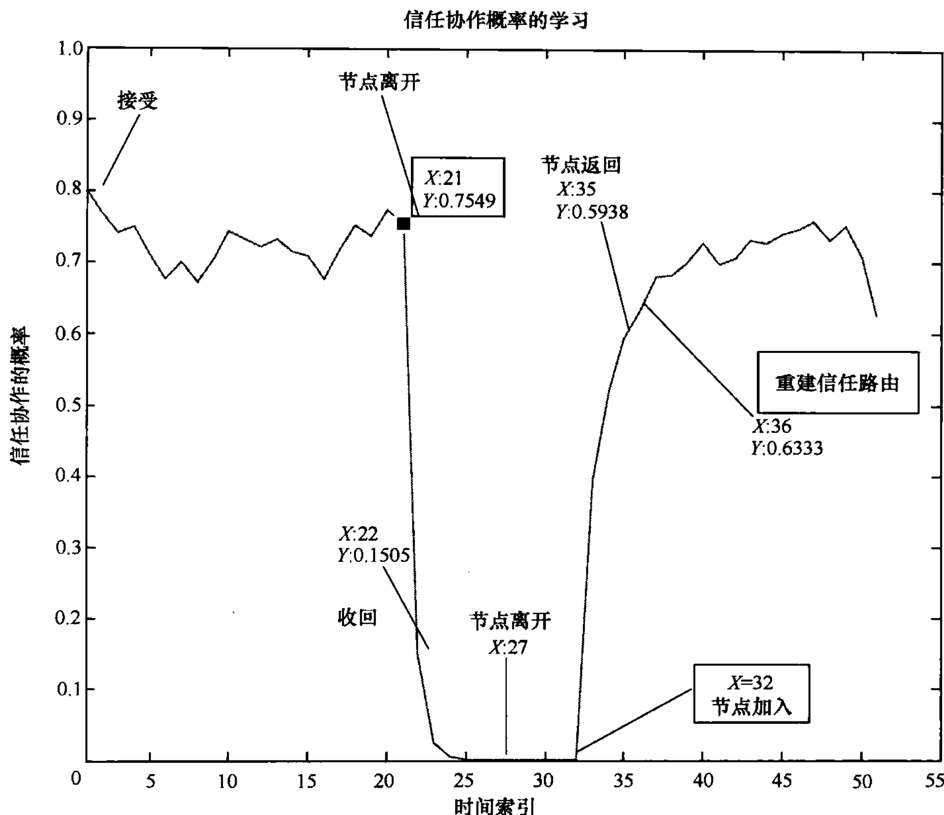


图 10-8　在繁忙传输密度中节点离开和加入网络

命题 10.10：（衰减因子重要性的拇指法则）引理 10.4 的学习算法在轻闲的传输密度下缓慢地响应新的信任证据。在我们唯一能采取的适应这种情况的方法是保持衰减因子尽可能小，并在随后较长的时间间隔中作出判决。

命题 10.11：（初值问题的拇指法则）初始值应尽快衰减，因为它反映了过去的信息。这违背了学习算法的基本概念，它应该试图回应最新信任的证据。我们可以用自适应衰减因子根据实际的网络条件去适当地和合适地“忘记”初始值。

10.3.2.3　节点行为的变化

在这个实验中，我们认为信任协作的概率趋向于更好的或者更糟的表现。认

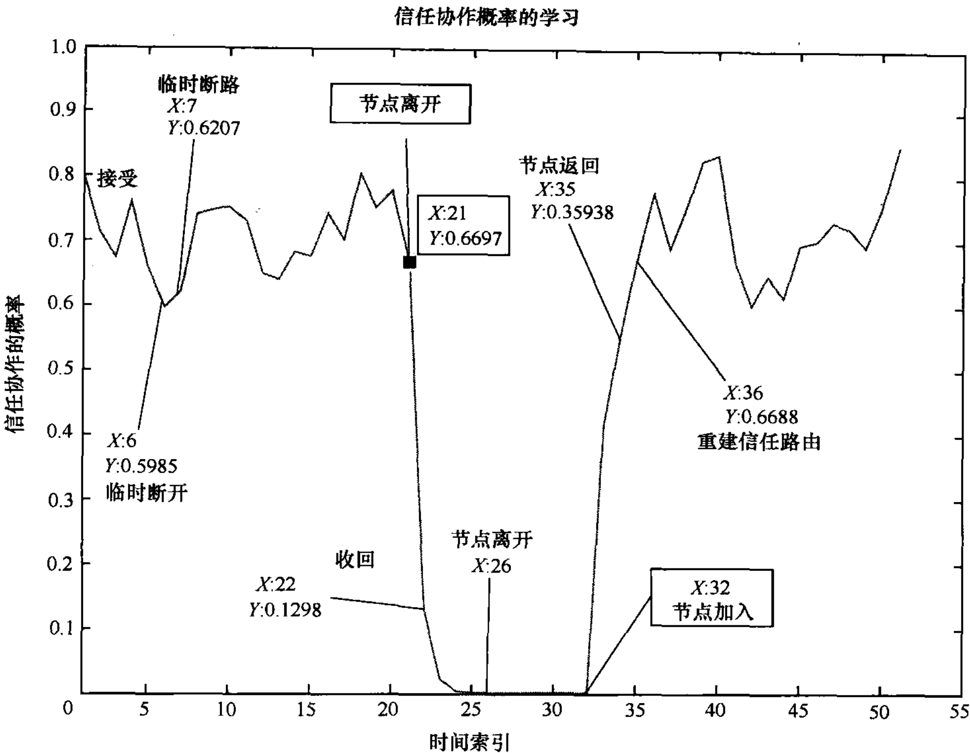


图 10-9 在轻闲传输密度下节点离开和加入网络

知无线电网络的节点可能招致一个恶劣的信道条件，因此，改变了它们的信任协作的行为。学习算法应该能够分析可能出现的暂时断开，并作出进一步的判决。在图 10-10 中，我们展示了节点改变其行为，在时间 $t=16$ 信任协作概率从 $p=0.9$ 到 $p=0.7$，然后在时间 $t=36$ 概率改变到 $p=0.9$，当 $k_1=0.2$，$k_2=0.5$ 时。学习算法检测到了在时间 $t=20$ 的变化，虽然信任协作的概率仍比信任阈值大，$p=0.7>0.6=\gamma_i$。原因在于衰减因子更多地比我们奖励良好行为而惩罚了不良行为。由于 $k_1=0.2<k_2=0.5$，在时间 $t=16\sim20$，学习算法检测到概率的预测的下降，然后在时间 $t=20$ 收回信任路由。从时间 $t=16$ 到时间 $t=35$，学习算法稳定地抓住了从节点转发的数据包，虽然预测概率相比于 $t=1\sim15$ 有明显下降。这可能有许多原因，在时间 $t=0.9$ 和 $t=0.7$ 之间的变化可能来自于信道条件。学习算法在时间 $t=37$ 将立即赶上概率从 $p=0.9$ 到 $p=0.7$，当信道状况好转，信任路由将会重建。在图 10-11 中，除了在每个时段的传输密度是 30，我们再次重复这个实验。我们注意到，在一个轻松的传输网络中预测曲线触发了行为变化中的干扰，我们标记这种现象为“临时断开”，如图 10-11 所示。这并不表明

一个信任路由的收回，而是代表了在信任阈值间的不稳定的振荡。但是，在捕捉行为的改变时，学习算法仍然行之有效，例如在时间 $t=16$ 和 $t=36$，最后检测的是信任协作的重建。

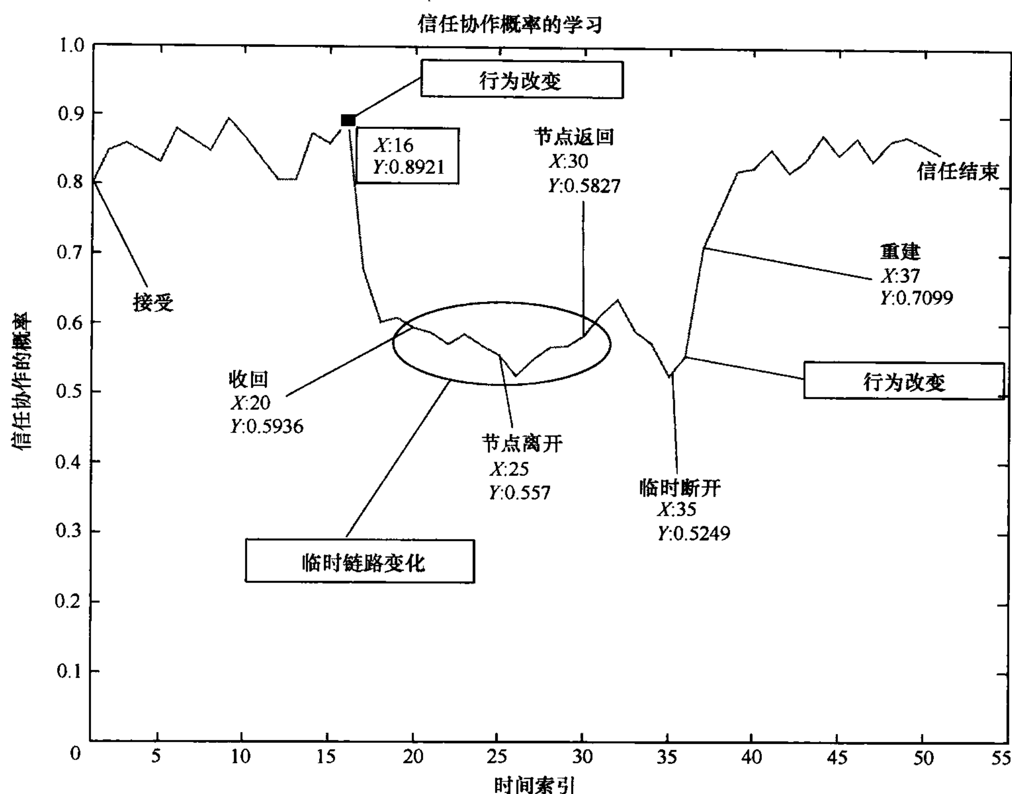


图 10-10　在轻闲的传输密度下节点改变行为

10.3.2.4　故意降低传输

学习算法不仅要解决网络中的动态频繁中断，也要面对一些特殊情况，如故意的降低传输。在这种情况下，节点不丢弃所有的传输。相反，它们减少一些传输的固定部分。如果丢掉的部分包含网络运行的重要参数，这可能导致整个认知无线电网络的严重损害。在图 10-12 中，我们展示在了在时间 $t=11$ 恶意用户的概率预测的显著下降和节点的通常行为。通过体现故意数据包丢弃的更大的衰减因子，恶意的用户行为被惩罚了。

上述例子描述了通过学习算法信任的自适应允许了认知无线电网络的期望操作。

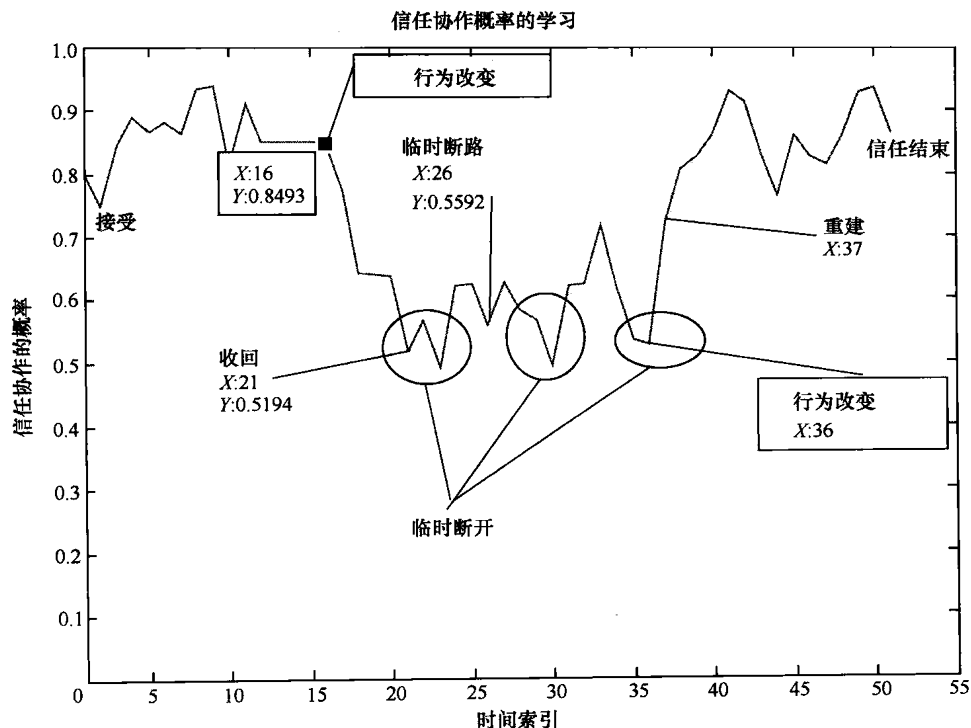


图 10-11　在轻闲传输密度下节点改变行为

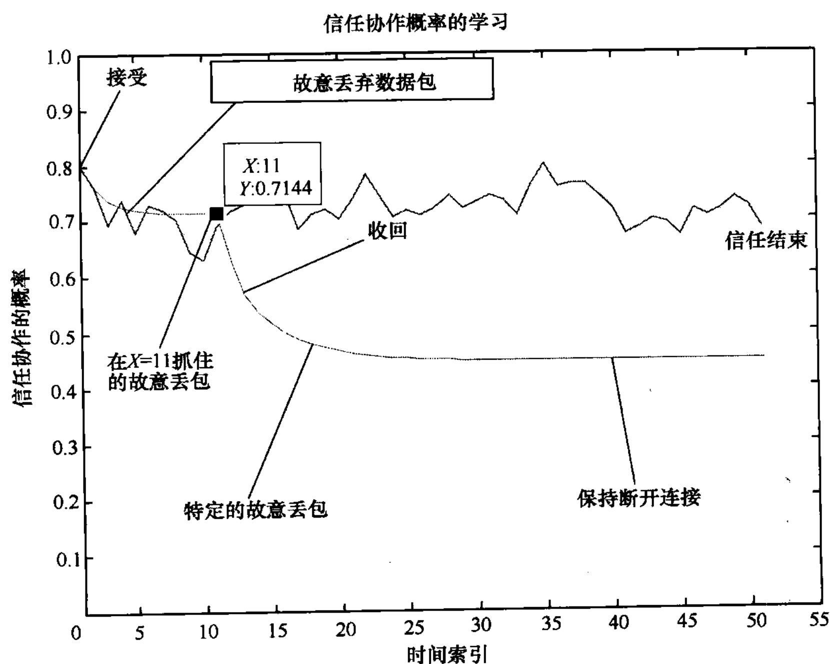


图 10-12　节点故意降低了传输的主要部分

10.4　CRN 中的安全机制

如第 9 章所述，CRN 可能构建于不同层次的信任和安全机制。前面章节已经描述了信任机制，在这一章中我们将会讨论安全机制。由于无线网络中的安全性已经得到广泛研究，因此我们着重讨论 CRN 安全性面临的新挑战。

10.4.1　蜂窝数据网络中的安全特点

我们首先讨论蜂窝网络中的安全性，蜂窝网络的系统框图如图 10-13 所示。UE 表示用户设备，RAN 表示无线电接入网络，CN 表示核心网络。蜂窝数据网络中安全服务的主要目标是保护信息和资源，以免受到攻击或者网络错误行为的影响。与其他无线数据网络例如 WLAN 相比，蜂窝数据网络的安全性具有如下特点：

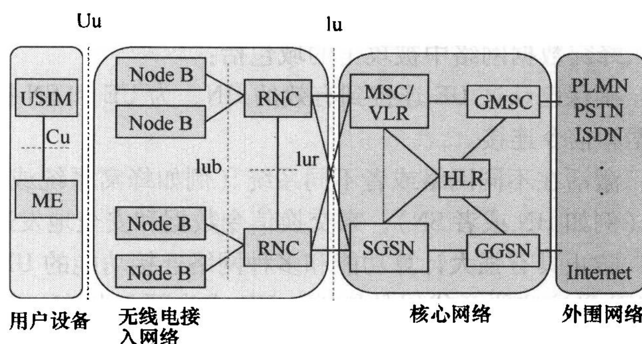


图 10-13　公众陆地移动网络（PLMN）的网络单元

■ 基于数据包的计费机制：这就是蜂窝数据网络中的安全性如此重要的原因。蜂窝数据网络向用户提供服务，并且基于用户收发的数据包的累计量计费。但是恶意用户为了减低资费或者增加服务量会绕过合法的认证、授权或者会计机制。因此，蜂窝数据网络中对用户的收费不能绝对基于用户的收发数据量，没有用户愿意为其他用户导致的欺诈数据付费。

■ 基于集中式网络架构的身份认证：集中式网络架构利用集中式用户数据库（归属用户站点，HSS）记录所有用户的身份和用户信息。国际移动用户识别码（IMSI）通常被用作用户的主要身份识别。为了防止恶意使用服务，网络通过挑战-响应机制来认证用户身份。挑战-响应机制给 UE 和 SN 提供会话钥匙，UE 和 SN 在会话钥匙的保护下建立连接。由于受益于用户和网络共享身份和预定义的钥匙信息，与现有的其他钥匙建立机制相比，挑战-响应机制更加简单。

■ 控制和数据传输分离：为了保持数据服务的 QoS，蜂窝数据网络通过隔离数据传输和控制传输来防止带内信令。但是由于有限的无线链路带宽，这种分离机制导致控制信道很容易过载。与传统方法相比，传输类似量的数据需要更多的控制信息来建立数据承载。在数据传输完成后，还需要交互大量信令信息来释放数据传输占用的无线电资源。这种建立和释放过程导致过高的信令负荷，使控制面很容易受过载攻击。

■ 分层网络结构：目前蜂窝数据网络的分层网络结构被认为有利于一些关键系统功能，例如位置管理、切换管理和功率控制。但是，这种结构增加了许多信令信息，并延长了信令时延，因此降低了整体性能。更糟糕的是，过重的信令负荷增加了系统被过载攻击的可能性。

为了讨论抵抗蜂窝数据网络中的攻击的方法，我们首先定义攻击：攻击通过利用一种方法或者系统脆弱性来违反某一特定域的安全属性。基于上述定义，我们将蜂窝数据网络的攻击分成两类：域和影响。我们通过分层的方式来介绍跟攻击有关的信息，首先介绍攻击所属的域，然后是攻击的影响，最后讨论导致攻击的系统脆弱性。蜂窝数据网络中被攻击的域包括：

■ 接入域：确保已认证 UE 连接到有效的 HN，为 UE 和 SN 提供保密的身份钥匙来保护数据和信令连接。

■ 网络域：激活在不同网络或者不同系统（例如蜂窝系统或者无线局域网）中的网络实体（例如 HN 或者 SN），来交换信令数据和安全地发送消息。

■ 设备域：防止具有强大计算功能和多种网络连接功能的 UE 被蠕虫或者病毒侵入。一旦 UE 受这些恶意代码的影响，UE 成为僵尸状态，并被用来将这些恶意代码扩散到其他用户或者攻击网络实体。

接下来我们介绍安全特性，并列举违反这些安全特性的攻击基本类型。由于一种攻击可能导致多种后果，为了简化，我们只考虑攻击的直接影响。任何一种已知的攻击都可以被分解成几个子组件，每个子组件可被归入如下的攻击技术：

■ 认证：蜂窝网络的一个主要优点是集中认证特性。认证机制通过用户识别来判别用户是否是所声称的用户，并确保只有合法用户才能使用网络。

● 会话攻击：恶意用户通过伪装成合法的基站来攻击已经建立的会话。

● 信息重播：攻击者可以截取加密的信息，随后重发给用户，导致用户无法判断接受的包是否是应该接收的包。

● 中间人攻击：一个攻击者可以位于蜂窝电话和接入站之间，截取它们之间的信息并改动。

■ 授权：允许已认证用户接入网络，并使用服务。为了实现网络接入控制，一些网络单元是必不可少的，例如缓冲。

● 缓冲溢出：为了导致缓冲溢出，攻击者通过将特意设计的代码植入与分

配的缓冲相邻的内存空间来破坏数据，缓冲溢出后，攻击者可以执行只有管理者具有权限的命令，并非法增加接入特许。

■ 保密：保护信息或者数据文件只能被已授权的用户接收，而不会暴露给其他用户或者被其他用户获取内容。保密通常可以通过加密实现。

● 窃听：如果无线链路中的信息没有加密，攻击者可以窃听并截取敏感的通信例如保密通话、敏感文档等。

■ 完整性：保证信息或数据文件在传输过程中不被恶意用户更改。完整性通常可以通过单向 Hash 实现。

● 信息伪造：如果通信信道不安全，攻击者可以截取双向信息并修改内容，而不被用户发现。

■ 不可否认性：确保实体不可否认其发出的信息。由于收发包都需要计费，不可否认性在蜂窝数据网络中尤其重要。蜂窝数据网络可以通过复杂的账单机制来达到不可否认性。

■ 可用性：保证期望的网络服务在有需求时即可获得。

● 信道干扰：攻击者可能通过干扰无线信道来阻止任何合法用户接入网络。

● 服务拒绝（DoS）：攻击者通过给网络发送过多的数据引起网络过载，导致用户无法接入网络。蜂窝网络过重的信令负荷和控制过程使其很容易受到此类攻击。

● 分布式服务拒绝（DDoS）：分布式服务拒绝类似于服务拒绝，但是分布式服务拒绝利用大量用户发送数据冲击网络带宽和资源。由于智能手机的广泛流行，攻击者很容易将病毒、蠕虫或者木马植入到智能手机中，从而控制大量用户发起此类攻击。

10.4.2　CRN 中的安全困境

与传统无线网络相比，CRN 面临新的安全问题。目前 CR/SDR 团体关注的焦点是通过预防式安全措施来保证无线电软件的安全下载，以及无线电软件安装成功后防止无线电软件被篡改。但是，在 CRN 中预防性安全错误不能起到完全保护作用，这是因为 CRN 更容易受干扰和自私基站的错误行为的影响。通过对用户的认证可以防止非授权用户使用网络。识别终端的常用方法是发射机验证过程，此过程通过用户位置验证和信号能量检测来完成。

CRN 中可能发起的 DoS 攻击通过阻止其他站点接入信道以达到提高自身吞吐量的目标。DoS 攻击导致其他用户不能获得网络服务并阻断已有的服务。在 CRN 中，站点通过交换 MAC 控制帧来竞争信道。由于缺乏基站，信道协商过程以分布式的方式进行或者通过频谱管理器实现。错误发送的控制帧会导致服务拒绝和自私的错误行为。这些问题可以通过 MAC 层的认证机制解决。

现在我们扩展前一节提到的安全顾虑：

■ 服务拒绝：在多信道环境中，高负载会引起频繁的控制数据包的交互，从而阻塞控制信道。从安全的角度考虑，如果攻击者能成功地阻塞控制信道，就能阻止信道协商和分配过程，从而引起服务拒绝。

■ 自私行为：在信道协商过程中，自私的站点会尝试利用更多的资源，从而提高自身的性能。信道协商需要利用频谱检测结果，而信道共享的公平性取决于相互竞争的站点的协作。但是自私的站点可能会隐藏那些可用的数据信道，以达到独享信道的目的。

■ 模拟主用户攻击：CR 使用空闲的拥有执照的频谱，在没有空闲的拥有执照的频谱时使用没有执照的频段。攻击者可以通过模拟主用户来干扰信道，使 CRN 只能使用没有执照的频段，因此限制 CRN 的网络吞吐量。遗憾的是，对于此种攻击目前还没有有效的解决方法。

■ 公共控制信道干扰：攻击者在控制信道发射周期脉冲。对控制信道的干扰可以阻止整个 CR 站点的可能通信。对于此类攻击的一种解决方案是采用基于超宽带 UWB 的公共控制信道。

■ 对频谱管理器的攻击：如果网络中只有一个频谱管理器分配频谱，只需要攻击频谱管理器就能导致所有 CR 站点不能相互通信。网络必须配置多个频谱管理器来分配频谱，频谱可用信息必须在 CRN 中有多个备份并分布于整个网络。防止此类攻击的另一种方案是在执照信道中使用导频信道。

■ 窃听：由于可以使用低于 UNI 的频段和 ISM，CR 的传输可能是长距离的。这种特性导致攻击者可以在不被应急服务可视的条件下收集数据。因此必须在物理层使用复杂的加密算法。

■ 隐私：在没有用户允许的前提下不能泄露数据。为了不危害用户的隐私，必须有规则和政策来约束上下文信息的发布和共享。

■ 网络管理架构：网络管理架构一般都基于分层的结构；在 Ad hoc 无线网络中为了被动和主动地监测覆盖，将分层结构做了进一步扩展。

■ 政策管理：在 CRN 中，政策管理的实现基于 CRN 管理机构发布更新概况信息。CRN 管理机构通过与用户的交互制定政策，并在 CRN 中传播政策更新信息。

10.4.3 CRN 中维护用户隐私的要求和挑战

除了安全性，无线网络必须保护用户隐私，这在例如 CRN 这样的异构无线网络尤其重要。CRN 中的隐私具有如下要求：

■ 维持信息隐私：即在没有用户批准时预防向其他服务或应用泄露跟用户直接相关的信息。

■ 维持上下文隐私：即预防泄露跟上下文有关的信息。这里上下文指用户使用服务相关的场景信息（例如目前设备参数）和可以提取用户信息的间接信息。

■ 维持位置隐私：即拒绝向攻击者提供设备当前和过去的位置信息，并预防可连接性。

■ 在特定场景下保持用户可识别参数的匿名：即保持用户不可识别的状态。匿名机制可以保证位置隐私，这是因为只要用户是匿名的，攻击者无法得知用户身份，从而保护位置隐私。

匿名机制必须在保护用户身份或者其他可识别身份的信息不被滥用的同时，允许用户使用网络服务。一种提供匿名的方法是使用标签名字/假名。对 CRN 而言，这要求使用用户身份的其他方式或者与用户信息相关的隐私，在必要的情况下包括使用匿名。匿名机制的弊端是如何管理匿名以及匿名的有效时间。

一种解决方案是基于能力的隐私保护方法。在这种方案中，政策管理器包含用户描述、上下文、场景信息、设备和服务。已知上下文的政策管理、隐私保护和匿名机制对授权用户透明并且是非突出的，即在所有情况下给予用户控制权，并且应该能方便地基于提供的模板创建和更新用户概况、安全和隐私政策、场景和政策。为了保持用户的匿名性，必须隔离用户身份相关的参数（名字、地址、身份证、银行账号等）和任何基于上下文的信息（位置、IP 地址、时间、出现、服务类型等）。

用户必须具有能够选择不同等级的用户隐私，并在保持用户隐私不受影响的前提下协商网络。保护用户隐私的政策包括：

- 请求最少必须的用户敏感数据来运行服务（数据最小原则）；
- 只在声称的时间段后保持数据；
- 在没有提前知会用户并取得用户许可的条件下不能任何第三方泄露数据；
- 明确表明向用户提供的位置和上下文信息的可信度；
- 明确表明为保证网络可信度的透明政策。

如下是上述建议的已知上下文的隐私保护机制的主要特点：

● 保护任何用户认为是隐私的敏感信息以及保护力度，即用户决定如何保护敏感信息和匿名性以及位置隐私。对所有低层次敏感的数据抽象也属于保护机制，数据抽象首先通过迅速过滤和默认设置来处理。

● 向用户隐藏系统的复杂性，使用描述性概况和政策、角色、场景和上下文来向设备报告用户的隐私决策。

● 基于规则的私有数据的接入可以帮助将决策委托给设备，在为设备预留私有参数方面采取行动。

● 上下文属性改变的任何时候，隐私保护机制重新评估整体隐私状态，并

基于预设定的规则作出响应。

10.4.4 CRN 中安全机制的实现

主用户的识别对准许接入和资源、服务调度至关重要，因此为了保证能够成功识别和认证主用户，必须使用主用户信号不可遗忘的特征。这就要求在结构上紧密结合 CR 识别和网络识别。简而言之，只使用一套用户证书更受欢迎。这里的解决方案是基于 802.1x 的 CRN 准入控制（CRNAC）。对用户识别而言，身份管理是必须的。身份管理指通过一系列工具和过程来管理组成用户数字身份的信息元素的生命周期。身份管理具有如下功能：

- 识别系统中交互的实体，认证每一个实体，应用政策、账单和审计说明的授权原则。
- 根据隐私管理规则给其他系统实体或者第三方提供实体相关信息（例如设备能力、设备特性、用户喜好或者角色）。

每一个 CR 站点需要激活 802.1x。当用户认证被请求时，CR 站点创建两个承载业务的虚拟端口。一个端口用于承载控制业务，另一个端口用于承载数据业务。在默认情况下，承担数据的端口没有激活，只有承载控制业务（无线可扩展认证协议，EAPOW）的端口被激活，并且控制端口在认证没有完成之前不会传输数据业务。图 10-14 所示为 802.1x 接入控制模型。

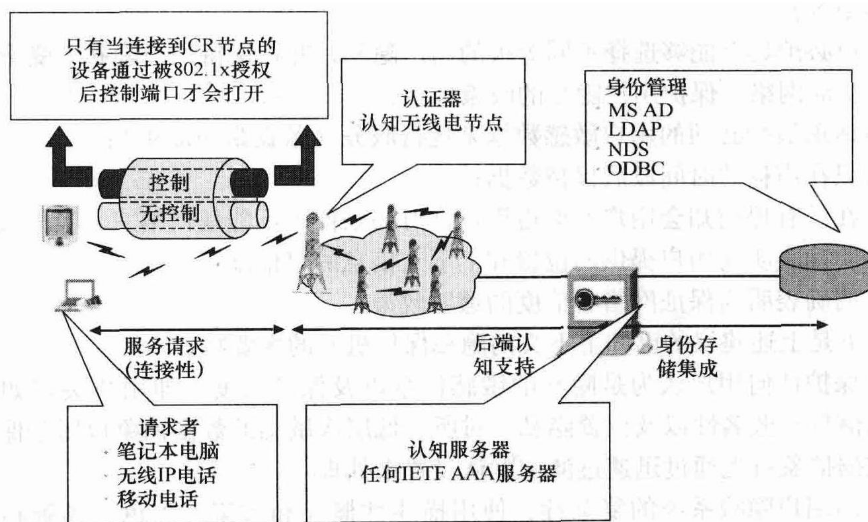


图 10-14 802.1x 接入控制模型

接下来需要一个安全角色或者清理服务器。为了更好的管理和控制安全系统的接入，安全角色是必须的，安全角色同时可加强政策管理工具的可用性。安全

角色与适宜的接入政策紧密相关，这是因为这些接入政策被应用于 CR 站点的接入控制列表。例如，政策决策点需要随时能够接入这些关联信息。当要求任何时间任何地点都能获得服务时，这种信息是必须的，这种需求给安全性带来巨大挑战。

为了达到如下目的，完整性和安全处理关联性是必须的。

- 避免可能危害系统的错误委派权利和准许用户特权。
- 当系统不能利用相关的管理和证书时，避免拒绝具有权限和特权的用户使用服务。

图 10-15 描述了 CRN 的安全框架结构。这种安全结果通过基于 802.1x 和 AAA 服务器的相互认知实现，并与第三层网络安全特性相结合，例如防火墙、虚拟私有网络、入侵检测和保护机制。

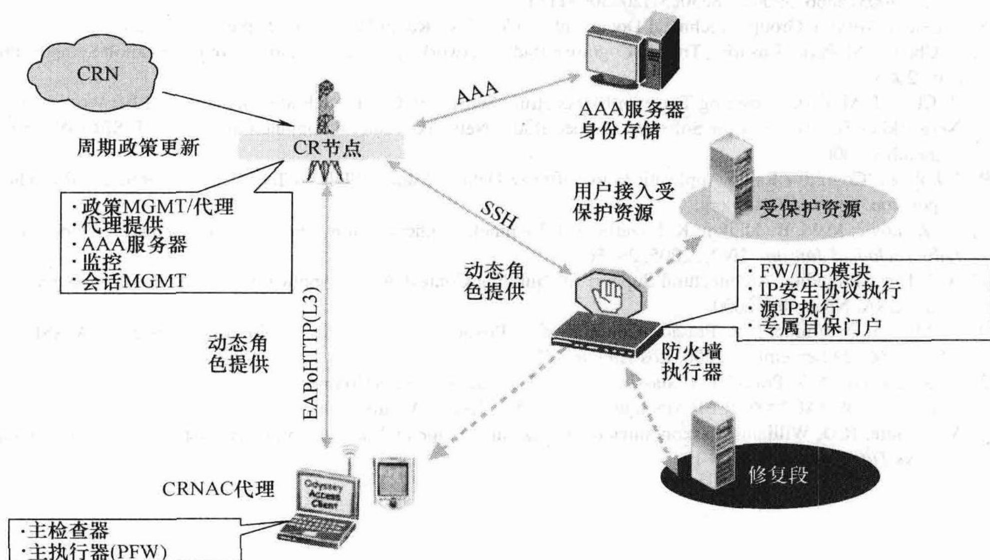


图 10-15　安全认知网络接入控制结构

参 考 文 献

[1] J. Mitola, III, G.Q. Maguire. 'Cognitive Radio: Making Software Radios More Personal', *IEEE Personal Communications*, **6**(4), 1999, 13–18.

[2] J. Mitola, III, *Cognitive Radio Architecture*, John Wiley & Sons, Inc., New Jersey, 2006.

[3] S. Haykin, 'Cognitive Radio: Brain-empowered Wireless Communications', *IEEE Journal on Selected Areas in Communications*, **23**(2), 2005, 201–220.

[4] K.C. Chen, L.H. Kung, D. Shiung, R. Prasad, S. Chen, 'Self-Organizing Terminal Architecture for Cognitive Radio Networks', *Proceeding Wireless Personal Multimedia Communications Conference, Jaipur, India, 3–6 December*, 2007.

[5] K.C. Chen, Y.C. Peng, N. Prasad, Y.C. Liang, S. Sun, 'Cognitive Radio Network Architecture: Part I – General Structure', *Proceedings of the ACM International Conference on Ubiquitous Information Management and Communication, Seoul*, 2008.

[6] K.C. Chen, Y.C. Peng, N. Prasad, Y.C. Liang, S. Sun, 'Cognitive Radio Network Architecture: Part II – Trusted Network Layer Structure', *Proceedings of the ACM International Conference on Ubiquitous Information Management and Communication, Seoul*, 2008.

[7] W. Rudin, *Functional Analysis*, McGraw-Hill, 1991.

[8] G. Theodorakopoulos, J.S. Baras, 'On Trust Models and Trust Evaluation Metrics for Ad Hoc Networks', *IEEE Journal on Selected Areas in Communications*, **24**(2), 2006, 318–328, 2006.

[9] H.C. Tijms, *Stochastic Models*, John Wiley & Sons, Inc., New Jersey, 1994.

[10] M. Puterman, *Markov Decision Process*, John Wiley & Sons, Inc., New Jersey, 1994.

[11] V. Poor, *Introduction to Detection and Estimation*, 2nd edition. Springer, 1994.

[12] Y.C. Ho, R.C.K. Lee, 'A Bayesian Approach to Problems in Stochastic Estimation and Control', *IEEE Trans. On Automatic Control*, October 1964, 333–339.

[13] R.P.S. Mahler, 'Multitarget Bayes Filtering via First-Order Multitarget Moments', *IEEE Trans. On Aerospace and Electronic Systems*, **39**(4), 2003, 1152–1178.

[14] S. Thrun, W. Burgard, D. Fox, *Probabilistic Robots*, MIT Press, Cambridge, MA, 2005.

[15] Bob Mayer, Patrick Kelly, 'Enhancing Wireless LAN Security with Cognitive Radios,' http://www.eetasia.com/ART_8800375666_590626_8e30c51f200509.HTM.

[16] Wireless Working Group, 'Technical Document on Cognitive Radio Networks', September 15, 2006.

[17] R. Chen, J.-M. Park, 'Ensuring Trust in Cognitive Radio Networks', Wireless Personal Communication Symposium, June 2006.

[18] R. Chen, J.-M. Park, 'Ensuring Trustworthy Spectrum Sensing in Cognitive Radio Networks', IEEE Workshop on Networking Technologies for Software Defined Radio Networks (held in conjunction with IEEE SECON 2006), September 2006.

[19] J. Polson, 'Cognitive Radio Applications in Software Defined Radio', SDR 04 Technical Conference and Product Exposition. www.sdrforum.org.

[20] M. Zivkovic, M.M. Buddhikot, K. Lagerberg, J. Bemmel, 'Authentication Across Heterogeneous Networks', *Bell Labs Technical Journal*, **10**(2), 2005, 39–56.

[21] A.K. Dey, 'Providing Architectural Support for Building Context-Aware Applications', PhD thesis, Georgia Inst. Tech., USA, November 2000.

[22] A. Mitseva, M. Imine, N.R. Prasad, 'Context-Aware Privacy Protection with Profile Management', WMASH'06, ACM 2006, 29 September, 2006, Los Angeles, CA.

[23] D. Kyriazanos, N.R. Prasad, G.I. Stassinopoulos, 'Ubiquitous Access Control and Policy Management in Personal Networks', IWUAC2006, IEEE Mobiqutous 2006, San Jose, CA, July 2006.

[24] V.M. Gure, R.D. Williams, 'Taxonomies of Attacks and Vulnerabilities in Computer Systems', *IEEE Commun. Surveys Tuts.*, **10**(1), 2008, 6–19.

第 11 章　认知无线电网络的频谱管理

认知无线电用户可以接入被主用户占据的需要授权的频谱，还可以通过一些宽带接入技术使用免授权许可频段。因此，认知无线电网络的运行类型可以划分为授权频带运行、免授权频带运行和混合多无线电共存三类。

■ 授权频带运行：授权频带主要是被主网络占用。因此在这种场景中，认知无线电网络主要是检测主用户。信道容量主要取决于临近主用户带来的干扰。此外，当在认知无线电用户占据的频带内出现主用户时，认知无线电用户应该放弃该频带，并立即转移到其他可用频段上。

■ 免授权频带运行：当没有主用户出现时，认知无线电用户与其他系统具有同样的接入频带的权利。因此，认知无线电用户需要使用成熟的频谱共享方法来竞争使用免授权频带。

■ 多无线电共存：在同一个地理区域，通过灵活的射频单元，认知无线电可以接入多个频带以及多个共存的无线电系统。如在第 5 章介绍的，需要在授权频带和免授权频带之间使用自组织协调器来协调工作。

在以上三类认知无线电网络运行方式中，认知无线电可以通过下列可能的方式接入认知无线电网络：

■ 认知无线电网络接入：认知无线电用户可以在授权频带和免授权频带接入认知无线电基站。由于所有的消息交互都是在认知无线电网络中进行，所以认知无线电网络的频谱共享策略与主网络相互独立。

■ 认知无线电 Ad hoc 接入：CR 用户可以与其他 CR 用户通过 Ad hoc 连接在授权和免授权频段上进行通信。

■ 主网络接入：CR 用户也可以通过授权频段接入主基站。与其他接入类型不同的是，CR 用户需要使用自适应媒体接入控制（MAC）协议，从而使用不同的接入技术在多个主网络之间进行漫游。注意，在这里提到的主网络可以是任何主系统（PS），也可以是现有的有线/无线通信基础设施架构或骨干网。

认知无线电网络需要解决与主系统/网络的共存，以及满足多种服务质量需求的问题。因此，CRN 频谱管理功能需要解决下列设计上的难点[1]：

■ 干扰避免：CRN 中的 CR 节点应该避免给主系统带来干扰。

■ QoS 保障：为了找到合适的工作频带，CRN 应该考虑动态和异构频谱环境，来支持需要保障服务质量的通信。

■ 无缝通信：无论主用户是否出现，CRN 应该支持无缝通信。

频谱管理过程由以下四个主要步骤组成[1]：

■ 频谱感知：CR 用户可以分配使用频谱中未被占用的部分。因此，这些用户应该监控可用的频段，获得可用频段的信息，并检测频谱空洞。在本书的第 7 章和 9.4 节中详细介绍了频谱感知技术。

■ 频谱判决：根据频谱可用性，CR 用户可以分配到一个信道。这种分配不仅仅是根据频谱可用性信息，也是根据内部（也可能是外部）策略判决出来的。第 5 章和第 8 章介绍了一些频谱判决技术的场景。

■ 频谱共享：可能有多个 CR 用户需要接入共享频谱，因此 CRN 接入应该在多个用户之间协调，来避免多个 CR 用户在重叠的频段上发生冲突。

■ 频谱移动性：CR 节点可以认为是频段上的漫游移动用户。因此，如果某频段被主用户占用，那么 CR 通信链路应该立即切换到另一个空闲的频段（或者是通过合适的方法来共享频段）。

CRN 的频谱管理功能，包括频谱感知、接入、网络层功能、QoS 等功能，可以归纳为图 11-1。

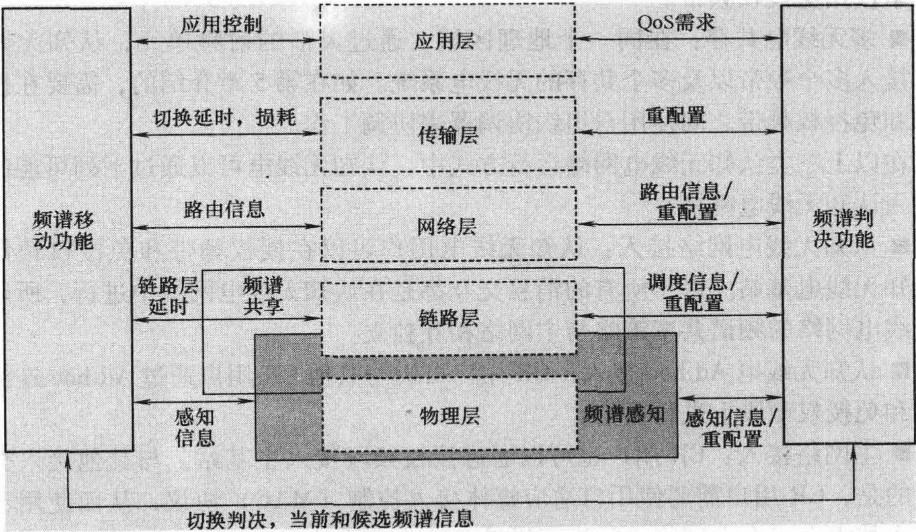


图 11-1 CRN 管理功能（文献 [1] 中的图 3）

在下面的章节中将介绍频谱管理中的频谱共享和移动性。

11.1 频谱共享

有效的频谱共享可能是实现 CRN 频谱管理的第一步。这是一个完全新的问

题，我们从同构 CRN 网络开始分析，也就是由 N 个节点组成的 Ad hoc 网络[2]。在这些节点中，有 L 个单播通信会话集。$s(l)$ 和 $d(l)$ 表示速率为 $r(l)$ 的会话 $(l \in L)$ 的源节点和目的节点。

我们还可以用 M_i 表示节点 $i \in \mathbb{N}$ 可用的频带集合，其中 $\mathbb{N}$ 表示 CRN 中的节点集合。$M = \cup_{i \in \mathbb{N}} M_i$ 是 CRN 中所有可用频带的集合。$|M| = M$ 表示 CRN 中所有可用频带的数量。$M_{ij} = M_i \cap M_j$ 是在 $\mathbb{N}$ 中节点 i 和节点 j 之间链路 (i, j) 的可用频带集。频带 $m \in M$ 的带宽为 W_m，并且该带宽可能分为 K_m 个子带，从而更灵活地分配资源，各个子带的带宽为 $u_{m,k} W_m$，其中

$$\sum_{k=1}^{K_m} u_{m,k} = 1$$

为了研究干扰，我们使用第 6 章在 $\mathbb{N}$ 中距离为 $d_{i,j}$ 的节点 i 和节点 j 的信道增益为

$$h_{i,j} = c_\alpha d_{ij}^{-\alpha}$$

CR-Tx 的功率谱密度（PSD）是 Q。在通过无线电环境传播后，如果接收到的功率谱密度大于门限 Q_T，那么就能成功接收。另一方面，如果接收到的干扰功率谱密度大于 Q_I，那么干扰将不可忽略。因此，传输范围是 $R_T = \sqrt[\alpha]{c_\alpha Q / Q_T}$，类似地，干扰范围为 $R_I = \sqrt[\alpha]{c_\alpha Q / Q_I}$。在一般情况下，$Q_I < Q_T$，等效为 $R_T < R_I$。

现在我们的问题是为一个节点分配用于传输和接收的子频带。设计可行的调度方法必须保证节点间没有干扰（或者干扰在可容忍范围内）。假设频带 m 可以用于节点 i 和节点 j。我们使用下面的指示函数（类似于媒体占据指示符）：

$$1_{ij}^{m,k} = \begin{cases} 1, & \text{如果节点 } i \text{ 和节点 } j \text{ 在子频带 } m_k \text{ 上通信} \\ 0, & \text{其他} \end{cases}$$

对节点 $i \in \mathbb{N}$ 和频带 $m \in M_i$，其他可以使用频带 m 的节点集合可以表示为

$$N_i^m = \{ j : d_{ij} \leqslant R_T,\ j \neq i,\ m \in M_i \}$$

因为节点 i 不能同时在一个子频带上向多个节点传输，即

$$\sum_{q \in N_i^m} 1_{iq}^{m,k} \leqslant 1 \tag{11-1}$$

同样，由于一般无线电通信中为了避免自干扰和信道泄露，节点 i 不能在一个子频带上同时传输和接收。用数学式表示为，当节点 i 发射时，对节点 j

$$1_{ij}^{m,k} + \sum_{q \in N_j^m} 1_{jq}^{m,k} \leqslant 1$$

换言之，如果 $1_{ij}^{m,k} = 1$，节点 j 不能使用该频带用于传输。如果 $1_{ij}^{m,k} = 0$，节点 j 可以使用该频带用于接收节点 $q \in N_j^m$ 的信息。

为了表示可能的干扰，我们用 P_j^m 表示在频带 m 上干扰节点 j 的节点集合，也就是

$$P_i^m = \{p : d_{pj} \leqslant R_I,\ p \neq i,\ N_p^m \neq \phi\}$$

因此，对于 $p \in P_i^m$，$p \neq i$，

$$1_{ij}^{m,k} + \sum_{q \in N_p^m} 1_{pq}^{m,k} \leqslant 1$$

在上面的等式中，如果 $1_{ij}^{m,k} = 0$，将会有两个节点对节点 j 产生干扰，但是这些节点远在干扰范围之外，因此仍然可能使用该频带用于传输。为了进一步简化等式，令

$$1_i^m = \{p : d_{pj} \leqslant R_I,\ N_p^m \neq \phi\}$$

实际上等效于

$$1_i^m = \begin{cases} P_i^m \cup \{j\},\ N_p^m \neq \phi \\ P_i^m,\text{其他} \end{cases}$$

限制条件可以表示为

$$1_{ij}^{m,k} + \sum_{q \in N_p^m} 1_{pq}^{m,k} \leqslant 1 (p \in N_i^m,\ p \neq i)$$

在 CRN 中，CR 源节点可能有一些协作中继节点（也包括中继路径）来把数据传输到目的节点，以实现更可靠的传输或者更大的累积带宽，从而构成多跳网络。通过把数据业务通过不同的路径和不同的子频带传输可以更好地实现频谱共享。我们用 $f_{ij}(j)$ 表示链路 (i, j) 上属于会话 l 的数据速率，其中 $l \in \mathbb{N}$，$j \in \cup_{m \in M_i} I_i^m$，$l \in L$。我们有如下三种可能的流平衡等式：

如果节点 i 是会话 l 的源节点，也就是 $i = s(l)$，那么

$$\sum_{j \in N_i^m} f_{ij}(l) = r(l)$$

如果节点 i 是会话 l 的中继节点，也就是 $i \neq s(l)$，$i \neq d(l)$，那么

$$\sum_{j \in N_i^m, j \neq s(l)} f_{ij}(l) = \sum_{p \in N_i^m, p \neq s(l)} f_{pi}(l)$$

如果节点 i 是会话 l 的目的节点，也就是 $i = d(l)$，那么

$$\sum_{p \in N_i^m} f_{pi}(l) = r(l)$$

每条无线链路的会聚流速率不能超过相应的链路容量，也就是如果节点 i 在子频带 m_k 上向节点 j 发送数据包，链路 (i, j) 在子频带 m_k 上的容量为

$$c_{ij}^{m,k} = 1_{ij}^{m,k} u_{m,k} W_m \log_2\left(1 + \frac{h_{ij} Q}{\sigma_n^2}\right)$$

链路容量限制可以表示为

$$\sum_{l \in L, i \neq d(l), j \neq s(l)} f_{ij}(l) \leqslant \sum_{m \in M_{ij}} \sum_{k=1}^{K_m} c_{ij}^{m,k} = \sum_{m \in M_{ij}} \sum_{k=1}^{K_m} 1_{ij}^{m,k} u_{m,k} W_m \log_2\left(1 + \frac{h_{ij} Q}{\sigma_n^2}\right)$$

因此对 CRN 的频谱共享需要找到一个最佳的每个节点上的可用频带的分配，

调度子频带传输和接收，以及协作中继路径分配流的方案，从而最小化 CRN 中所需的总带宽。

命题 11.1：CRN 频谱共享是通过下面的条件找到 $1_{ij}^{m,k}$、$u_{m,k}$和$f_{ij}(l)$

$$\min \sum_{i \in \mathbb{N}} \sum_{m \in M_i} \sum_{j \in N_i^m} \sum_{k=1}^{K_m} 1_{ij}^{m,k} u_{m,k} W_m \tag{11-2}$$

并且满足下列限制

(i) $\sum_{k=1}^{K_m} u_{m,k} = 1(m \in M)$

(ii) $\sum_{q \in N_i^m} 1_{iq}^{m,k} \leqslant 1(i \in \mathbb{N}, m \in M_i, 1 \leqslant k \leqslant K_m)$

(iii) $1_{ij}^{m,k} + \sum_{q \in N_p^m} 1_{pq}^{m,k} \leqslant 1(i \in \mathbb{N}, m \in M_i, 1 \leqslant k \leqslant K_m, j \in N_i^m, p \in \Gamma_l^m, p \neq i)$

(iv) $\sum_{l \in L, i \neq d(l), j \neq s(l)} f_{ij}(l) \leqslant \sum_{m \in M_{ij}} \sum_{k=1}^{K_m} 1_{ij}^{m,k} u_{m,k} W_m \log_2\left(1 + \frac{h_{ij} Q}{\sigma_n^2}\right)(i \in \mathbb{N}, j \in N_i^m)$

(v) $\sum_{j \in N_i^m} f_{ij}(l) = r(l)(l \in L, i = s(l))$

(vi) $\sum_{j \in N_i^m, j \neq s(l)} f_{ij}(l) = \sum_{p \in N_i^m, p \neq s(l)} f_{pi}(l)(l \in L, i \neq s(l), i \neq d(l))$

这个最优化问题属于混合整数非线性规划问题，一般都是 NP 难题。因此对于 CRN 频谱共享问题，只能找到具有较低复杂度的次优解。最后，要注意上面的最优化问题是基于频域得到的，但是同样的原理可以根据合适的物理层限制条件适用于时域或者逻辑信道。

11.2　频谱定价

到目前为止，我们讨论的 CRN 没有为主系统带来任何好处，并且 CRN 会在某些时间段干扰主系统。因此，需要给主系统用户提供合适的收入和激励措施来允许灵活的使用频谱，从而更有利于实现实际的 CRN。为了实现有效的动态频谱共享，需要为频谱/系统运营商和频谱使用者建立一个经济模型，从而最大化激励（也就是收入和利润）和用户满意度。我们一般把频谱拥有者当做运营商。频谱使用者可以是主系统用户，也可以是次要系统用户（也就是认知无线电用户）。主系统用户的优先级一般要高于次要（CR）系统用户，这取决于 CRN 运行的具体规则策略。

定价和资源分配是紧密联系在一起的。由于服务提供商（也就是运营商）想最大化自身的收入，用户（主系统用户和次要系统用户）想最大化用户的满意度，这两方面一般用 QoS 性能和价格来衡量。在这个场景中有三方参与者：

运营商、主用户、次要（CR）用户。此外管理者可能是 CRN 中的第四方，代表着公共利益，通过最小化每个用户的代价来最大化频谱利用率，我们在之后的章节讨论这种情况。这三方有不同的效用函数和不同的优化自身利益的动机。运营商想最大化频谱利用率来提高收入/利润。当有未被分配的频带时，运营商可以分配和把频带用伺机使用的方式租赁给次要认知无线电用户。运营商可以获得额外的收入和利润。主系统和次要系统用户想用更小的成本代价获得更好的服务（更多带宽和更好的信道质量），而主系统用户可以从中减少自身的支出。由于三方各自有不同的效用函数，它们可能相互竞争或协作来获得更多利益。例如，当次要用户想获得更好的 QoS，这会导致在同一个频段上的主系统用户 QoS 降低。我们可以简单的定义三方的效用函数如下：

$$U_{\text{Operator}} = R_{\text{Primary}} + R_{\text{Second}} - \text{成本}$$

$$U_{\text{Primary}} = V(\text{频谱大小，Qos}) - \text{成本(付出)}$$

$$U_{\text{Secondary}} = V(\text{频谱大小，Qos}) - \text{成本(付出)}$$

U_{Operator}的成本来源于授权频谱的收入、用户的补偿等。Niyato 和 Hossam 在参考文献 [18] 中提出了多个主服务提供商（运营商）相互竞争，为次要 CR 用户提供频谱接入机会，以提高频谱利用率。通过定价方案，每个主服务提供商在主用户的 QoS 限制下最大化自身的利润。这种场景可以通过 Bertrand 博弈模型建模成由一些企业（运营商）和一个用户（次要 CR 用户）组成的寡头市场，如图 11-2 所示。

运营商和次要 CR 用户之间的交互可以归纳如下。由于与次要用户共享频谱导致主用户 QoS 下降可以认为是运营商的成本代价。运营商需要补偿主系统用户服务质量下降。在 Bertrand 模型中分析了系统参数对博弈的影响，例如频谱可持续性（表示在纳什均衡点时次要业务在提供不同的主服务和信道质量的运营频段上切换的能力）。使用分布式算法可以得到该博弈的最优解。由于该博弈的纳什均衡解不是最大化运营商的利益，因此提出了在运营商之间协商的方案来最大化运营商的利益。由于协商的方案并不是纳什均衡点，所以参与者（运营商）想偏离该方案来提高自己的利益。这种行为会损害其他参与者的利益。因此需要引入惩罚机制来强制主

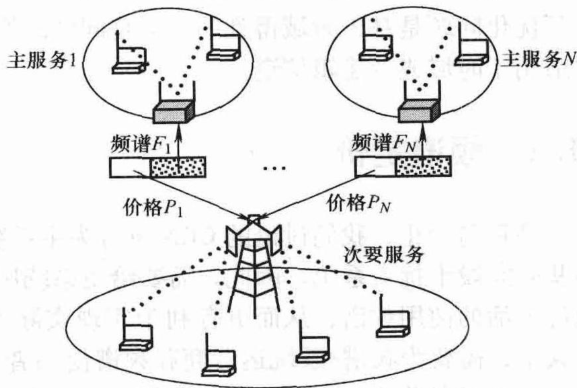


图 11-2 频谱共享的系统模型（见参考文献 [18]）

服务提供商考虑自己的长期利益，从而避免运营商的私自背离行为。如果由于采取协商机制带来的长期利益要比背离带来的短期利益高，那么运营商将会采取协商机制。

在博弈论中，使用效用函数来量化次要服务的频谱需求。那么对于主服务，可以定义向次要服务提供频谱接入的成本代价。该代价函数是根据本地连接服务质量下降定义的。一般用下面的二次效用函数作为次要效用函数：

$$u(b) = \sum_{i=1}^{N} b_i k_i^{(s)} - \frac{1}{2}\left(\sum_{i=1}^{N} b_i^2 + 2v\right) - \sum_{i=1}^{N} p_i b_i$$

式中，b 是由所有主服务共享的频带大小的集合，也就是 $b=\{b_1, b_2, \cdots, b_N\}$，$p_i$是主服务 i 提供的价格，$k_i^{(s)}$是次要用户使用的频谱效率。该效用函数通过参数 v 还考虑到了频谱的可替代性。为了得到频谱f_i的需求函数，将 $u(b)$ 对 b_i取微分，即

$$\frac{\partial u(b)}{\partial b_i} = 0 = k_i^{(s)} - b_i - v\sum_{j\neq i} b_j - p_i$$

我们可以通过求解上面的等式得到在给定所有主服务定价时的需求函数。频谱需求函数可以表示为

$$D_i = \frac{(k_i^{(s)} - p_i)(v(N-2)+1) - v\sum_{i\neq j}(k_i^{(s)} - p_i)}{(1-v)(v(N-1)+1)} \tag{11-3}$$

主用户 i 的收入是 $R_i = c_1 M_i$，成本代价是

$$C_i(b_i) = c_2 M_i\left(B_i^{\text{req}} - k_i^{(p)}\frac{W_i - b_i}{M_i}\right)^2$$

式中，c_1和 c_2分别表示收入和代价函数的权重。B_i^{req}是主连接所需的带宽。主用户 i 的利润是

$$P_i(p) = b_i p_i + R_i - C_i(b_i)$$

根据之前提到的系统模型，可以构建一个 Bertrand 博弈。博弈中的参与者是主服务。每个参与者采取的策略是频谱的单位价格（用 p_i表示）。每个参与者的回报是利润。博弈的最佳解是纳什均衡点。为了得到最佳解，需要解下面的等式。对所有 i，$\frac{\partial P_i(p)}{\partial p_i}=0$。如果我们想最大化主服务的总利润，我们需要求解下面的等式：

$$\frac{\partial \sum_{j=1}^{N} P_j(p)}{\partial p_i} = 0 \tag{11-4}$$

对主服务 i，令 P_i^0、P_i^n、P_i^d分别表示最佳定价、纳什均衡定价和偏离纳什均衡定价带来的利润。那么，对于一直保持协商机制的场景，主服务 i 的长期利润

可以表示如下：

$$P_i^0+\delta_i P_i^0+\delta_i^2 P_i^0+\cdots=\frac{P_i^0}{1-\delta_i}$$

如果一个主服务偏离了最佳定价，该主服务会在第一个阶段获得偏离的利益，而在之后的阶段，主服务将获得纳什均衡点的利益。因此，主服务 i 的长期利润可以表示如下：

$$P_i^d+\delta_i P_i^n+\delta_i^2 P_i^n+\ldots=P_i^d+\frac{P_i^n}{1-\delta_i}$$

如果由于通过协商获得的长期利润比偏离协商获得的利润要高，那么将维持协商，也就是：

$$\frac{P_i^0}{1-\delta_i}\geqslant P_i^d+\frac{P_i^n}{1-\delta_i}\Rightarrow\delta\geqslant\frac{P_i^d-P_i^0}{P_i^d-P_i^n}$$

本文构建了重复博弈来分析这种情况。我们分析了这种场景下运营商、主用户和次要用户的交互，如图 11-3 所示。

这种场景考虑了多个运营商之间的均衡、在竞争环境下竞价的稳定性、次要用户在频谱可持续性时的频谱需求函数以及采用策略的稳定性。但是，并没有对主用户和次要用户的服务质量有任何保障。只是通过经济机制来降低 QoS 和限制用户（包括主用户和次要用户）来补偿主用户。并不在所有用户之间进行接入控制，或者考虑终端用户之间的频谱分配和 QoS 需求。重复博弈可能需要多次迭代才能达到稳定状态。基于不同服务模型的不同数量用户和用户的使用时间（主用户和次要用户）是影响频谱利用率的重要因素，因此也需要在建模时加以考虑。

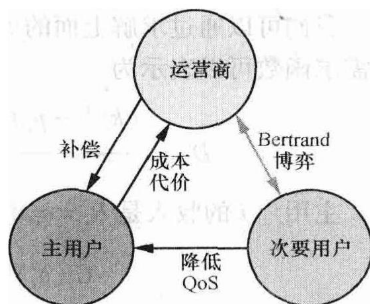


图 11-3 运营商、主用户和次要用户的交互

我们还进一步考虑在一个主用户和多个次要用户之间频谱共享的问题[19]，仍然使用寡头市场竞争和非协作式博弈来获得次要用户的频谱分配策略。静态博弈是假设次要用户能获得目前采用的策略和每个参与者的收益等信息，有限理性的次要用户可以根据之前采用策略的观察结果来迭代的动态调整策略（也就是动态博弈）。

图 11-4 所示的分析场景中包括一个主用户和 N 个次要 CR 用户。主用户想把部分频谱（b_i）共享给次要用户。在一般寡头市场的非协作博弈模型中，所有企业以产品质量进行竞争。在这种场景中，次要用户就是企业，每个用户想竞争频带大小（也就是产品质量）。次要用户的利润来自于主用户的收费和分配频谱

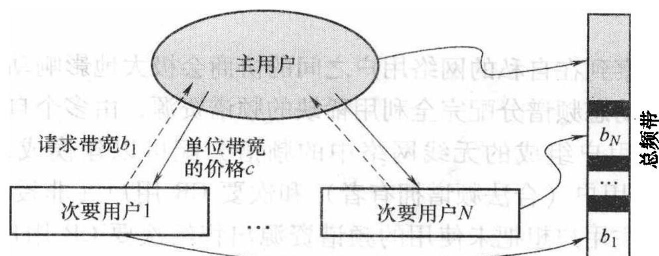


图 11-4　频谱共享的系统模型[19]

获得的利益。在静态博弈模型中，博弈的参与者是次要用户，每个参与者的策略是请求和分配的频谱大小（用 b_i表示次要用户 i 的频谱大小），每个次要用户的收益函数表示利润。主用户的定价函数定义为

$$c(\boldsymbol{B}) = x + y\left(\sum_{b_j \in \boldsymbol{B}} b_j\right)^{\tau} \tag{11-5}$$

式中，x、y 和 τ 是非负常数，$\tau > 1$（因此定价函数是凸函数），$\boldsymbol{B}$ 表示所有次要用户的集合（也就是 $\boldsymbol{B} = \{b_1, \cdots, b_N\}$）。次要用户 i 的收入可以表示为 $r_i \times k_i \times b_i$，其中 r_i是次要用户 i 单位传输速率能获得的收益，k_i是次要用户 i 可以达到的传输速率。因此，次要用户 i 的收益可以表示为 $\pi_i(\boldsymbol{B}) = r_i \times k_i \times b_i - b_i c(\boldsymbol{B})$。在给定其他次要用户 b_j，$j \neq i$ 共享频谱大小时，次要用户 i 的最优响应可以定义为 $BR(\boldsymbol{B}_{-i}) = \mathrm{argmax}_{b_i} \pi_i(\boldsymbol{B}_{-i} \cup \{b_i\})$。数学上，可以获得纳什均衡的充分必要条件是

$$\frac{\partial \pi_i(\boldsymbol{B})}{\partial b_i} = r_i k_i - x - y\left(\sum_{b_j \in \boldsymbol{B}} b_j\right)^{\tau} - y b_i \tau \left(\sum_{b_j \in \boldsymbol{B}} b_j\right)^{\tau - 1} = 0 \tag{11-6}$$

与上面的静态博弈不同，动态博弈调整请求/分配频谱大小的方法可以建模成：

$$b_i(t+1) = Q(b_i(t)) = b_i(t) + \alpha_i b_i(t) \frac{\partial \pi_i(\boldsymbol{B})}{\partial b_i(t)}$$

式中，$b_i(t)$ 是在时间 t 时分配的频谱大小，α_i是次要用户的调整速率参数（也就是学习速率）。

我们可以从图 11-5 中看到这种场景中主用户和次要用户之间的交互，该图主要集中在授权和非授权用户需要和谐共存。主用户是否可以自己决定价格或者共享频带而不需要通过运营商的控制，这是一个重要的问题，

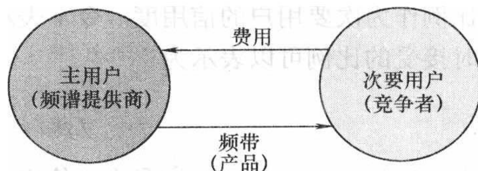


图 11-5　主用户和次要用户的交互

因为这样会给其他主用户带来干扰。然而，这种场景可以确保有效地提高频谱利

用率。

我们可以观察到在自私的网络用户之间的协商会极大地影响动态频谱共享的效率。为了通过动态频谱分配完全利用稀缺的频谱资源，由多个自私的合法频谱拥有者和非授权用户组成的无线网络中的频谱分配可以建模成多阶段动态博弈[20]，在多个主用户（合法频谱拥有者）和次要 CR 用户（非授权用户）共存的网络场景中，主用户想把未使用的频谱资源出售给次要 CR 用户来获取利益。例如，主用户可以是连接到核心网络的频谱经纪人，也可以是归属于具有 CR 技术基站的次要用户；主用户还可以是网状网的接入点和次要用户移动设备。这种竞价博弈可以建模成拍卖博弈。

主用户或次要用户之间串通竞标是一种主要的欺诈行为。具体来说，竞标者相互串通在拍卖中以很低的价格（或者更高的价格）获胜。这种行为会导致偏离竞争均衡（CE）：也就是想购买的买家的数量等于想出手的卖家数量时的价格。为了对抗用户串通，提出了基于竞价的对抗串通动态频谱分配方案来优化频谱效率，不仅仅激励自私用户参与拍卖频谱，还避免了用户之间串通竞价。避免串通的办法是设置一个最佳的反向价格。反向价格是根据串通用户数量而定的。但是串通用户数量可能是未知的（也就是不完备信息）。参与者需要构建其他参与者未来可能采用的策略来辅助自己做出决策判断。

运营商预期从次要用户通过租赁频带（反向价格为 ϕ_{r,p_i}）获得的收入可以表示为

$$E_{V_i}[U_{p_i}(a_i^j,\phi_{r,p_i})] = (\phi_{r,p_i} - E[b_i])(F_{C_M^{-i}}(\phi_{r,p_i}) - F_{C_{M-1}^{-i}}(\phi_{r,p_i})) + \int_{\phi_{r,p_i}}^{Mb} (z - E[b_i])f_{C_M^{-i}}(z)\mathrm{d}z \tag{11-7}$$

假设上面的等式存在一个最大点，那么最优的反向价格是 ϕ_{r,p_i}^*，并且在有 C_{M-1}^{-i} 和 C_M^{-i} 分布的先验信息时，满足上面等式的一阶导数条件：

$$F_{C_M^{-i}}(\phi_{r,p_i}^*) - F_{C_{M-1}^{-i}}(\phi_{r,p_i}^*) - F_{C_M^{-i}}(\phi_{r,p_i}^*)f_{C_{M-1}^{-i}}(\phi_{r,p_i}^*) = 0 \tag{11-8}$$

有时运营商知道用户之间串通竞价，因此如何在考虑限制的条件下获得最优反向价格仍然是一个问题。因此建议把次要用户在不同的价格水平可以接受的竞价比例作为次要用户的信用度。令 y 表示次要用户的竞价价格。次要用户在 y 价格时接受的比例可以表示为

$$\tilde{r}_S(y) = \frac{\eta_A(y)}{\eta(y)}$$

式中，$\eta_A(y)$ 和 $\eta(y)$ 分别是在 y 价格时被接受竞价的数量。如果很多买家都同时参与到竞价时，一般可以精确估计 $\tilde{r}_S(y)$。我们可以观察到，如果竞价 $\tilde{y} < y$ 被拒绝，那么竞价 y 也会被拒绝；如果竞价 $\tilde{y} > y$ 被接受，那么竞价 y 也会被接

受。为了找到次要用户的信用度，对每个在 y 价格时的潜在竞价，定义

$$\overset{\cup}{r}_S(y)=\begin{cases}1, & y=0\\ \dfrac{\sum\limits_{w\geqslant x}\eta_A(w)}{\sum\limits_{w\geqslant x}\eta_A(w)+\sum\limits_{w\leqslant x}\eta_R(w)}, & y\in(0,M_b)\\ 0, & y\geqslant w\end{cases}$$

式中，$\eta_R(w)$ 是以价格 w 竞价被拒绝的竞价数量，M_b是足够大的值，因此比 M_b 价格高的竞价肯定会被拒绝。因此，上面的信用度函数可以表示为 C_M^{-i}的累积分布函数。

这种场景只考虑了运营商和次要用户的共存，或者主用户和次要用户的共存，而没有同时考虑所有三方博弈参与者。这种场景的目标是提高频谱利用率和最大化总效用，运营商和次要用户或主用户和次要用户之间的交互如图 11-6 所示。

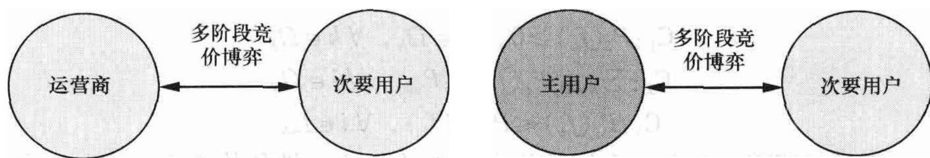


图 11-6 运营商、主用户和次要用户的交互

通过多阶段动态博弈，在主用户和次要用户之间通过双向定价过程协商动态频谱分配来最大化主用户和次要用户的效用，并且避免自私用户的串通行为。设计了一个分布式避免串通的动态定价方案和最优预留价格来有效地分配频谱，并且避免用户串通。通过为每个用户引入信用函数可以极大地降低判决时的竞价负荷。此外，考虑到可能有用户串通，可以使用纳什协商方案来达到提出方案的性能下限。这种方法的限制在于次要用户不能根据它们对频谱的需求进行竞价。

允许 CR 用户能动态接入频谱，并且不干扰影响有频谱许可的主用户，提高频谱利用效率最关键的挑战是如何设计有效的 MAC 机制来自适应的根据周围的环境在 CR 用户之间分配传输功率和频带。可以通过分布式竞价方法设计一种新颖的联合功率/信道分配方案，从而建模成非协作博弈。本节提出了一种基于定价的迭代注水（PIWF）算法来允许 CR 用户达到纳什均衡点，并且 CR 用户分布式地重复协商各自的最佳传输功率和频带。同时还提出了一种 MAC 协议来执行无线资源管理功能。

该算法主要考虑在认知无线电用户与主用户共存的伺机认知无线电网络，此外还需要考虑认知用户之间的交互。博弈模型的参与者是次要用户，对应可以采取

的策略是不同的传输功率。博弈的目标是最大化各自的传输速率，使用频谱竞价来提高次要用户总的传输速率。首先提出了在频谱管理中 MAC 协议执行的功率限制。

每个非协作 CR 用户都试图最大化自己的传输速率。这种贪婪机制可以建模成一个博弈：$\theta=\{\Omega, P, \{U_i\}\}$，$\Omega=\{1, 2..., N\}$（CR 用户的链路），其中 $\boldsymbol{P}=\boldsymbol{P}_1\times\boldsymbol{P}_2\times\cdots\times\boldsymbol{P}_N$是功率 $\boldsymbol{P}_i$的策略空间，U_i：$\boldsymbol{P}\rightarrow\Re$ 是参与者 i 的效用函数。参与者是 CR 用户。为了最大化传输速率，博弈的目标是纳什均衡点，虽然这种均衡点不是全局最优总传输速率。也就是说，博弈达到的纳什均衡点可能不是 Pareto 最优。实际上，我们的目的是最大化所有用户的总效用，用 $\max_p\sum_{i\in\Omega_N}U_i(P_i,P_{-i})$ 表示。为了得到能靠近 Pareto 最优的纳什均衡点，使用竞价机制来激励自私的用户协同工作。具有竞价的新效用函数定义为 $\widetilde{U}_i(P_i,P_{-i})=\sum_{K\in\Omega_K}\widetilde{u}_i(P_i(f_k))$。该博弈用数学式定义如下：

$$\begin{aligned}&\max_{P_i}\widetilde{U}_i(P_i,\ P_{-i}),\ \ \forall i\in\Omega_N\\&C_1:P_i(f_k)\geqslant 0,\forall i\in\Omega_N,\ \ \forall k\in\Omega_K\\&C_2:\textstyle\sum_{K\in\Omega_K}P_i(f_k)\leqslant P_{\max},\ \ \forall i\in\Omega_N\\&C_3:P_i(f_k)\leqslant P_{\text{mask}}(f_k),\ \ \forall i\in\Omega_N\end{aligned}\tag{11-9}$$

如果对于博弈 θ 存在一个纳什均衡点，并且这个纳什均衡点是 Pareto 最优，那么线性竞价函数因子是

$$\lambda_i(f_k)^{\text{opt}}=\sum_{j\in NBR_i}\frac{h_{jj}(f_k)P_j(f_k)h_{ij}(f_k)}{M_j(f_k)(M_j(f_k)+h_{jj}(f_k)P_j(f_k))}\tag{11-10}$$

线性竞价函数可以通过拉格朗日乘子来得到。用户 i 的拉格朗日乘子可以表示为

$$\begin{aligned}J_i&=\sum_{k\in\Omega_K}\widetilde{u}_i(P_i(f_k))+\sum_{k\in\Omega_K}\alpha_{i,k}P_i(f_k)-\beta_i\Big(\sum_{k\in\Omega_K}P_i(f_k)-P_{\max}(f_k)\Big)\\&=\sum_{k\in\Omega_K}\widetilde{u}_i(P_i(f_k))-\lambda_i(f_k)P_i(f_k))+\sum_{k\in\Omega_K}\alpha_{i,k}P_i(f_k)-\beta_i\Big(\sum_{k\in\Omega_K}P_i(f_k)-P_{\max}(f_k)\Big)\end{aligned}\tag{11-11}$$

竞价因子 $\lambda_i(f_k)$ 越高，就表示用户 i 在信道 k 上使用较大的传输功率。为达到纳什均衡点提出了两种算法：串行 PIWF 和并行 PIWF。每个独立的 CR 用户 i，首先在所有的信道根据博弈方程式调节自己的竞价因子 $\lambda_i(f_k)$，然后通过在所有信道上测量总的噪声加上干扰 M_i来确定自己的最优策略，也就是最优的信道/功率/速率。用户 i 的最优反应是在限制 $C_1\sim C_3$ 下最大化自己的效用函数。在 CRN 中对每个用户重复上面的步骤。如果用户是根据一定的顺序依次确定自己的最优反应判决，则有下面的串行 PIWF 算法。

算法 11.1　串行 PIWF 算法

0：初始化 $P_i(f_k)=0$，$\forall i\in\Omega_N$，$k\in\Omega_K$，初始化迭代计数 $l=0$。

1：重复迭代；

2：$l=l+1$；

3：for $i=1$ to N 用户 do

4：for $k=1$ to N 信道 do

5：估计总的干扰与噪声的和 $M_i(f_k)$；

6：计算竞价因子 $\lambda_i(f_k)$；使用步骤 5；

7：使用接收到的控制数据的信号功率估计信道增益 $h_{ii}(f_k)$。

8：end for

10：在选择的信道上使用下面的功率传输

$$\boldsymbol{P}_i^{(l)}=BR_i(\boldsymbol{P}_{-i})=[\boldsymbol{P}_1^{(l)},\ \cdots,\ \boldsymbol{P}_{i-1}^{(l)},\ \boldsymbol{P}_{i+1}^{(l-1)},\ \cdots,\ \boldsymbol{P}_N^{(l-1)}]$$

11：end for

12：对所有 $i\in\Omega_N$直到 $l>L_{\max}$或 $(\|\boldsymbol{P}_i^{(l)}-\boldsymbol{P}_i^{(l-1)}\|/\|\boldsymbol{P}_i^{(l-1)}\|)\leqslant\varepsilon$。

反之，如果用户是同时做出各自的最优反应决策，那么就按下面的并行 PIWF算法。

算法 11.2　并行 PIWF 算法

0：初始化 $P_i(f_k)=0$，$\forall i\in\Omega_N$，$k\in\Omega_K$，初始化迭代计数 $l=0$。

1：重复迭代；

2：$l=l+1$；

3：for $i=1$ to N 用户 do

4：for $k=1$ to N 信道 do

5：估计总的干扰与噪声的和 $M_i(f_k)$；

6：计算竞价因子 $\lambda_i(f_k)$；使用步骤 5；

7：使用接收到的控制数据的信号功率估计信道增益 $h_{ij}(f_k)$。

8：end for

9：$\boldsymbol{P}_i^{(l)}=BR_i(\boldsymbol{P}_{-i})=[\boldsymbol{P}_1^{(l)},\ \cdots,\ \boldsymbol{P}_{i-1}^{(l)},\ \boldsymbol{P}_{i+1}^{(l-1)},\ \cdots,\ \boldsymbol{P}_N^{(l-1)}]$

10：end for

11：for $i=1$ to N 用户 do

12：使用功率 $P_i^{(l)}$传输

13：end for

14：对所有 $i\in\Omega_N$直到 $l>L_{\max}$或 $(\|\boldsymbol{P}_i^{(l)}-\boldsymbol{P}_i^{(l-1)}\|/\|\boldsymbol{P}_i^{(l-1)}\|)\leqslant\varepsilon$。

上述两种 PIWF 算法都需要每个 CR 准确地估计系统参数，但是实际上很难估计系统参数。为了克服这种问题，提出了一种宽松的算法，每个 CR 用户只需要记录下自己最近选择的策略和其他用户选择的策略。把算法 11.1 和算法 11.2

中的最优反应函数用下面的函数替换：

$$\boldsymbol{P}_i^{(l)} = w\boldsymbol{P}_i^{l-1} + (1-w)\boldsymbol{BR}_i(\boldsymbol{P}_1^l, \cdots, \boldsymbol{P}_{i-1}^l, \boldsymbol{P}_{i+1}^{l-1}, \cdots, \boldsymbol{P}_N^{l-1})$$

就可以实现宽松的串行和并行 PIWF 算法，其中 w 可以理解成内存因子。

通过交易的方法，在参考文献［15］中提出了一种可以在 CRN 中满足所有参与者需求的频谱管理方法。

11.3 异构无线网络中的移动性管理

CRN 作为一种异构无线网络，很有可能是基于因特网协议（IP）的网络。目前几乎没有任何与 CRN 移动性管理相关的文献，在本节中我们主要介绍在异构无线网络中的移动 IP 管理。

为了实现对不同的网络系统透明，IP 协议用于不同网络互连的协议。基于 IP 互连的优点是隐藏了底层技术与高层技术的异构性。IP 也是一个广泛应用的通信协议，全面地为网络应用有效地提供了高效低成本的网络基础设施。IP 网预计将是 CRN 对应的核心骨干网。Ian F. Akyildiz、Shantidev Mahanty 和 Jiang Xie 在参考文献［27］中通过定义两个新实体：网络互操作代理（NIA）和互连网关（IG），提出了构建移动 IP 架构的方法。例如，如果有一个移动终端/节点从 GPRS 网络漫游到 WLAN 网络。如图 11-7 所示，IG 在 GPRS 网络中，连接网关 GPRS 支持节点（GGSN）和 NIA。在 WLAN 网络中，IG 连接了接入路由器（AR）和 NIA。IG 负责移动性管理、业务管理、鉴权和计费。NIA 由运营数据库、切换管理、鉴权和计费组成。很多管理单元例如移动性、业务、鉴权和计费都构建在 IG 中。在这些例子中，运营数据库和切换实体构建在 NIA 中。此外，在 NIA 的切换管理和 IG 的移动性管理中考虑了 QoS。为了实现 QoS，NIA 和 IG 都需要考虑调度功能。这种问题使得实现移动性管理更加复杂。

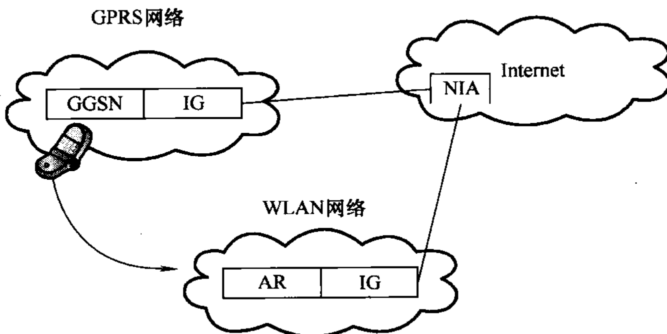


图 11-7　移动 IP 网络中基于 NIA 的架构

为了实现移动 IP 中的切换管理，这种架构需要有三个新的功能实体：本地代理（HA）、外地代理（FA）和移动终端（MT）或移动节点（MN）。这些实体的关系如图 11-8 所示。

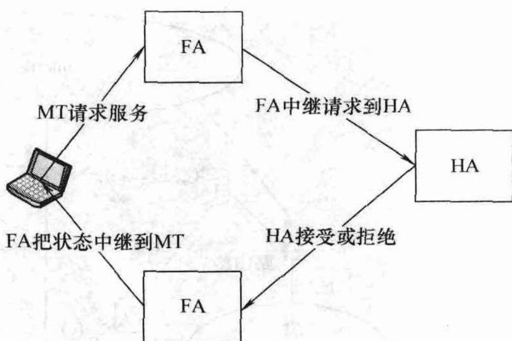


图 11-8　移动 IP 中三个新功能实体的关系

移动用户在两个网络域之间的移动称为宏移动；移动用户在一个网络域的两个子网之间移动称为微移动。这与之前我们提到的水平切换和垂直切换一样。宏移动可以总结如下：

■ 发现代理：MN 能够通过周期性地接收到每个 FA 广播的主动代理广播消息来检测自己是否移动到新的子网。MN 还可以发送代理请求消息来了解所有可能的移动代理。

■ 注册：当 MN 发现自己处于外地网络，MN 获得新的转交地址（CoA）。可以通过请求或监听 FA 广播（FA CoA）来得到 CoA，或者通过动态主机配置协议（DHCP）或点到点协议（PPP）来获取 CoA。MN 在 HA 处注册新的 CoA。然后 HA 把 MN 的 CoA 与 MN 的永久 IP 地址关联，更新移动绑定。

■ 路由和隧道：通信对端节点（CN）发送到 MN 的数据包通过 HA 进行解析。HA 封装数据包，并通过隧道发送到 MN 的 CoA。通过 FA 的 CoA，封装的数据包到达服务 MN 的 FA，然后 FA 解封装数据包，把数据转发到 MN，如图 11-9 中的步骤 a、b 和 c。通过临时配置的 CoA，封装的数据包到达 MN，然后 MN 解封装数据包。在图 11-9 中，隧道的终点是 MN 而不是 FA。

当 MN 从一个子网移动到另一个子网，切换过程如图 11-10 所示的下列步骤。这就解释了移动 IP 的切换管理。

■ 当 MN 进入新的子网时，获得新的 CoA。

■ MN 向 HA 注册新的 CoA。HA 建立起新的隧道，其终点是新的 CoA，并且移除旧 CoA 的隧道。

■ 一旦建立起新的隧道，HA 把数据包使用 MN 新的 CoA 通过隧道发送到 MN。

在切换中还要考虑其他因素，例如降低切换带来的服务延迟或者减少服务中断。为了保证服务质量，移动代理会在移动终端或用户切换到另一个（异构）无线网络之前请求预留资源。此外，较好的带宽预留策略可以有效地解决由于网络拥塞造成的服务中断，并且减少新生成的呼叫和切换呼叫的阻塞概率。CRN 中的微切换目前仍然是一个开放的问题。

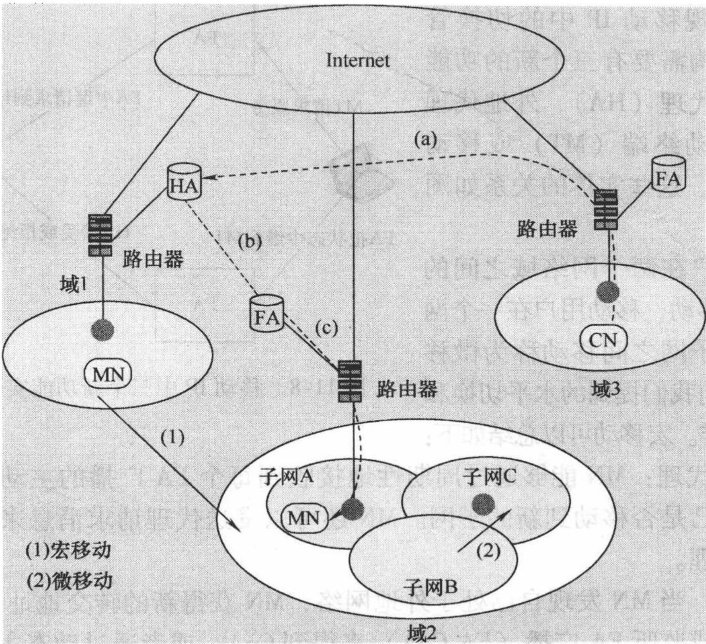


图 11-9 移动 IP 架构（见参考文献［28］）

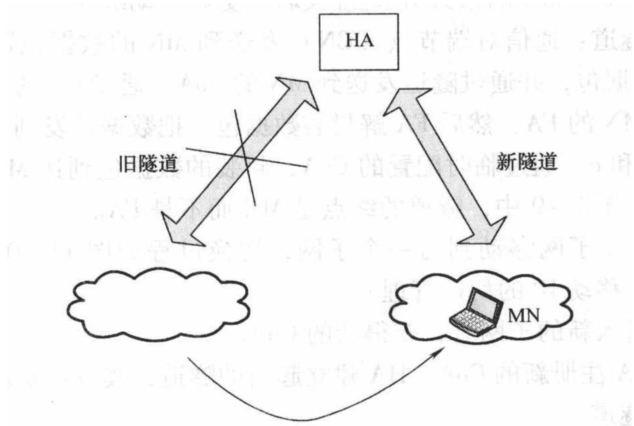


图 11-10 当 MN 移动到新子网，HA 建立起新隧道

11.4 管理问题和国际标准

CRN 不仅仅指一种技术，也涉及到在 11.2 节中介绍的经济因素。此外，CRN 还包括了频谱管理策略和正在进行的国际标准化工作。

“认知”这个词最初指代的是设备能够感知周围环境状况，并相应地调整自己的行为的能力。因此，认知无线电是基于人类理解、学习和适应环境的方法。J. Mitola Ⅲ定义的认知无线电如下：

认知无线电智能地识别出个人无线数字助理（PDA）、相关网络所需的无线电资源和相应的计算机到计算机通信，从而实现：

（a）智能地侦测用户的通信需求。

（b）并为这些需求提供最适合的无线电资源。

其他认知无线电的特征，包括观察、调整、归纳、学习、规划、判决和执行，如图 11-11 所示。目前终端的可重配置能力是用户设备独有的能力，在用户设备移动到另一个位置或者根据用户的不同需求来手动或自动变换和重新配置终端。CR 是一种智能无线电，可以通过感知环境和根据周围环境来调整自己的传输，对用户而言完全透明。

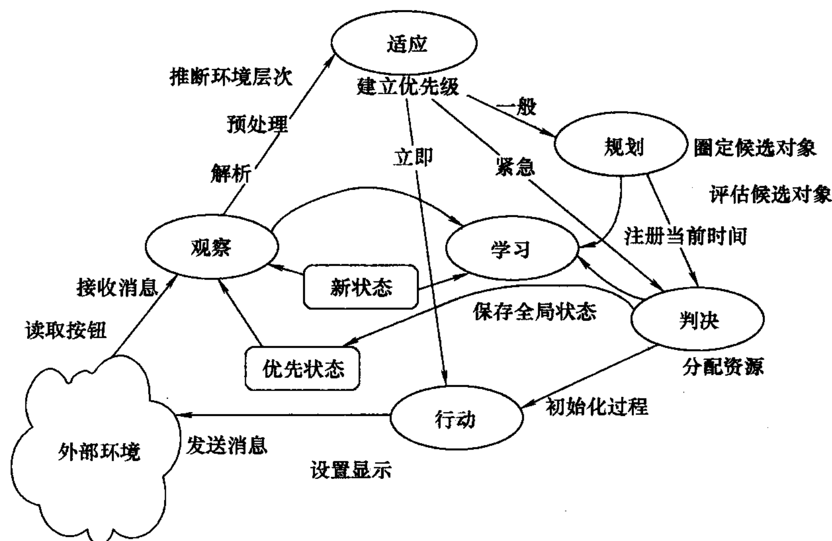


图 11-11　简化的认知无线电高层运行

在认知无线电技术实现之前，还需要解决很多问题。从潜在的技术问题到部署认知无线电的管理方面，所有这些问题都需要详细分析和研究，从而提供面向未来应用的灵活有效的无线电平台。在这些问题中，提高频谱利用率和保证对主用户干扰在可接受范围内是两个最重要的问题。

11.4.1　管理问题

认知无线电在管理方面最主要的应用是无线电资源管理（RRM）、频谱管理

和限制干扰。RRM 的主要目标是有效地使用有限的无线电资源。频谱管理则是通过频谱共享或动态频谱分配来有效地利用可用频谱。大量的异构设备具有不同的干扰功率，认知无线电最主要的问题是如何限制对授权系统的干扰功率在可接受范围内。

可重配置网络操作主要取决于有效管理可用资源：

1）更有效的利用可用频谱；

2）智能的网络规划过程；

3）管理属于不同无线电接入网络的固定分配的无线电频谱资源；

RRM 是比较复杂但是又必不可少的过程。RRM 包括动态管理资源（例如频谱），还包括动态在异构无线接入基础设施中的无线电接入技术（RAT）分配业务。因此，RRM 可以看成是频谱管理和联合无线电资源管理（JRRM）的合集。

在新的频谱管理中，某些无线电接入技术没有固定分配的频段。可以通过智能管理机制向这些无线电接入技术动态分配频带，从而最大化每个无线电接入技术的容量，并且减少对主系统的干扰。无线电资源不仅仅指无线电频谱，在实际无线电网络中还体现在每个移动用户的接入优先级、移动用户活动的时间段、信道编码、传输功率、连接模式等，需要在不同的时间尺度设计管理功能。JRRM 是用于在异构网间通信的控制机制，在异构网中移动终端和无线电网络由很多无线电接入技术共存组成，并且这些无线电接入网相互协作，但是频谱分配固定。这些机制同时工作在多个无线电接入技术上，此外必须有可重配置/多模终端的支持。

资源管理和认知无线电之间关系很紧密：认知无线电为终端提供了在任何时刻为用户实时确定最佳工作频段，以及提供用户所需服务的技术方法。

无线技术的日益发展和无线多媒体服务需求的日益增长导致了可用频谱稀缺的问题。多媒体服务相比现在的 3G 技术需要较大的带宽，而目前的频谱管理方法显然满足不了带宽的需求。与无线电频谱稀缺这样的传统观念不同，实际的研究中发现大多数已分配的频谱在特定的时间和位置都没有得到利用。动态频谱接入和动态分配频谱比实际频谱稀缺这一问题更加重要。因此，需要改变现在的频谱管理政策。

为了提高频谱利用率，在 2004 年美国联邦通信委员会（FCC）发布了规则制定建议通知（NPRM），提出了允许未授权用户/次要用户从授权频谱者临时借用频谱的可能性，只要不对主用户带来不可容忍的干扰。与传统把频谱分为两部分（授权和免授权频段）不同，动态频谱分配可以提供更灵活的协调频谱利用的方法。认知无线电可以感知频谱，在与授权频谱用户共存的同时重配置认知无线电设备的传输（功率等级、调制方案和协议）。

11.4.1.1 认知无线电的管理认证

认知无线电需要在管理认证方面使用统一的认证测试过程来确保认知设备能

够顺利工作。IEEE 1900 委员会最近成立了一个研究组来研究根据认知无线电的动态频谱接入规范正常工作的技术方面和标准化工作。频谱管理策略在不同的国家、不同的频谱拥有者、不同的时间和地理位置上差别很大。在任何一种工作模式下，为了保证频谱管理的连贯性，认知无线电需要把这些频谱策略作为认知库的一个组成部分。在美国先进防卫研究项目机构——neXt 代（DARPA XG）项目中正在进行把频谱策略用机器可以读取识别的语言表示的工作。本体网络语言（OWL）正是基于这个目的开发的语言。此外还需要一个管理框架和实施来监测、检测和停止频谱策略。以更安全的方法，例如数字认证或轻量级加密系统，来更新策略。可重配置软件不能修改安全模块，安全模块应该具有较强的防篡改能力。

11.4.1.2 并行传输

认知无线电技术的基础是填补空白的无线频段。但是研究者还在寻找更灵活和更有效的方法来同时或同频与主用户并行传输。在这种场景中，认知无线电不需要等待空闲信道来传输数据。但是在这种信道中传输可能有很多限制，例如两个用户可以同时传输的数据速率，当用户设备不是认知无线设备时需要比较设备参数，并且还要面对很多管理和工程上的问题。同时还必须进行测量来保证并行传输的次要用户不会影响主用户的传输。

11.4.1.3 认知无线电间的协作

工作在同一个频段上的认知无线电之间的协作可以减少检测时间，因此提高了整体灵活性。与完全不协作的方案相比，认知无线电之间的协作提高了灵活性。

11.4.1.4 干扰

认知无线电不仅仅造成了小区外干扰，还包括系统外干扰。很多异构设备具有不同的功率、占空比和传输路径损耗，这些都是干扰的来源。如果认知无线电能够靠近主用户，并且相应地调整自己的传输功率，那么就能抑制干扰。实际的认知无线电设备需要有足够的灵敏度来检测在阴影效应下的低信噪比信号。因此认知无线电主要考虑的是要保证对授权或合法的主系统的干扰在可接受范围内。最基本的问题是认知无线电系统是否能保证不对合法的用户带来有害的干扰，因为任何再可靠的频谱感知技术也会有出错的情况。更确切地说，认知无线电可能不能可靠的检测到主用户的存在，因此当主用户在使用无线电资源时认知无线电设备可能仍然会进行发送，此时就需要将第 7 章中相关的技术用于实际的频谱管理。

另一个考虑这个问题的角度是干扰可以用干扰温度限来度量。干扰温度定义在特定的地理区域和特定的频率。2003 年，FCC 提出了用于量化和管理干扰的干扰温度概念。通过接收功率，认知无线电可以测量现在的干扰环境，并且通过调整自己的发射参数来避免对授权用户带来有害干扰，从而使用空闲频率。因

此，提出了类似 CDMA 网络中的干扰温度多址接入（ITMA）用干扰温度来管理功率和带宽。这是一个新的适用于多用户 Ad hoc 无线通信网络的复用技术，它可以在不干扰已有信号的前提下共享使用授权频带。目前正在进行主用户和基于 ITMA 的次要频谱用户之间交互的研究。

11.4.1.5 信标概念

认知无线电网络中的频谱共享可以划分为下面两类：

● 垂直频谱共享：次要用户周期性搜寻空闲频带来传输数据。一旦感知到主用户的存在，次要用户将放弃工作频带，并切换到另一个空闲频带。

● 水平频谱共享：如果次要用户检测到某个频带上另一个次要用户，并且没有较好的其他工作频带，次要用户将与其他次要用户共享使用该频带。

如图 11-12 所示，垂直和水平频谱共享是认知无线电根据目前工作在同一频段上的无线电系统的管理状况决定的。水平共享对次要无线电系统而言更加灵活和自由。而垂直频谱共享的频谱协商则需要使用基于信标的方法。

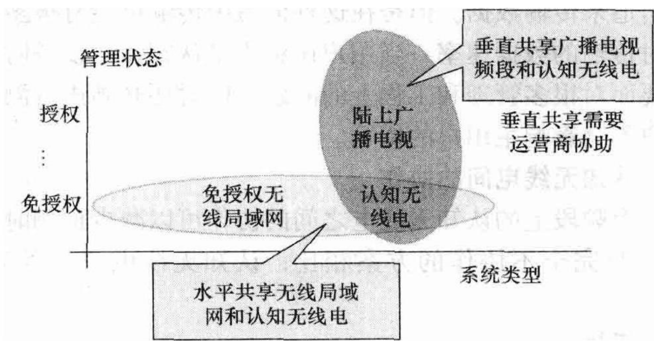


图 11-12 垂直共享和水平共享：取决于其他无线电系统的管理状态，CR 与其他类型的系统共享频谱

信标这一概念是为运营商提供动态频率分配的辅助信令机制。可以用信标来集中地决策协商利用次要频谱。允许（授予）频谱接入和拒绝频谱接入是主用户传输的两种信令机制或信标。当信标是独立使用时，这些信标可靠性有限。但是当信标是同时使用时，可以更有效地协商频谱接入。信标可以用不同的方法实现，例如以某一功率传输某种信号，并以某一灵敏度进行检测。检测授权信标和非拒绝信标只允许认知无线电使用频谱。同时检测到两种信标让认知无线电不能使用频谱。信标机制确保了更高的可靠性和更小的干扰。

11.4.2 国际标准

目前有一些正在进行制定的认知无线电相关的国际标准。世界无线研究论坛

（WWRF）有一个认知无线电的工作组，还有一个自组织网络的特别兴趣小组。IEEE 802.19 正在研究多无线电共存系统的技术方案，认知无线电技术无疑正是其中之一。

IEEE 802.22 工作组成立于 2004 年 11 月，主要研究方向是无线区域网（WRAN）。该工作组致力于统一基于认知无线电在 TV 广播频段免授权工作的空中接口（也就是 MAC 和 PHY 层协议）。该标准在认知无线电技术的发展中起着重要的作用，其目标是定义一个在任何管理体内可以管理的国际标准[5]。

IEEE P1900 标准组是在 2005 年第一季度由 IEEE 通信协会和 IEEE 电磁兼容协会共同建立的工作组。IEEE P1900 标准组的目的是开发支持下一代无线电和频谱管理新技术的标准。这个发起组也称为 SCC41——标准协调 41 委员会。其下属的单个工作组仍然称为 IEEE P1900. x。

在 IEEE P1900 标准组内有下面工作组：

- IEEE P1900. 1：频谱管理、策略定义无线电、自适应无线电和软件定义无线电的标准术语、定义和概念。
- IEEE P1900. 2：干扰和共存分析的操作规程建议。
- IEEE P1900. 3：软件定义无线电（SDR）软件模块一致性评估的操作规程建议。
- IEEE P1900. a：动态频谱接入无线电系统的管理依赖性和评估。目前是一个研究组，已经提交了项目授权请求（PAR），但是，最终决定把该研究组合并到 P1900. 3。
- IEEE P1900. 4：允许网络设备在异构无线接入网络中分布式最优化无线电资源使用的网络架构模块。

虽然认知无线电的标准化工作目前大多集中在链路层，但是随着因特网规模的日益扩大，预计认知无线电网络将在未来几年成为现实。

参考文献

[1] I.F. Akyildiz, W.-Y. Lee, M.C. Vuran, S. Mohanty, ‘A Survey of Spectrum Management of Cognitive Radio Networks’, *IEEE Communications Magazine*, April 2008, 40–48.

[2] Y.T. Hou, Y. Shi, ‘Spectrum Sharing for Multi-hop Networking with Cognitive Radio’, *IEEE Journal on Selected Areas in Communications*, **26**(1), 2008, 146–155.

[3] FCC, Cognitive Radio Workshop, 19 May, 2003, [Online]. Available: http://www.fcc.gov/searchtools.html.

[4] J. Mitola, Cognitive Radio: An Integrated Agent Architecture for Software Defined Radio, Ph.D. dissertation Royal Institute of Technology (KTH), 2000.

[5] C. Cordeiro, K. Challapali, D. Birru, S.N. Shankar, ‘IEEE 802.22: The First Worldwide Wireless Standard based on Cognitive Radios’, *IEEE Journal*, 2005.

[6] J. Mitola III, G.Q. Maguire, Jr., ‘Cognitive Radio: Making Software Radios More Personal’, *IEEE Personal Communications*, **6**(4), 1999, 13–18.

[7] J. von Neumann, O. Morgenstein, *Theory of Games and Economic Behavior*, Princeton University Press, New Jersey, 1947.

[8] D. Fudenberg, D.K. Levine, *The Theory of Learning in Games*, MIT Press, Cambridge, MA, 1999.
[9] J. Mitola III, 'Signal Processing Technology Challenges of Cognitive Radio'.
[10] 'Information on IEEE P1900 Purpose, Plans, Progress and Relevance to the work of ITU-R WP8F', May, 2005.
[11] P. Martigne, 'Overview of Some Standardization Activities on Cognitive Radio', ETSI Workshop on SDR/CR, February 2007.
[12] D. Cabric, I.D. O'Donnel, M.S.-W. Chen, R.W. Brodersen, 'Spectrum Sharing Radios', *IEEE Circuits and Systems Magazine*, Second Quarter 2006.
[13] NTIA, 'U.S. frequency allocation', [Online]. http://www.ntia.doc.gov/osmhome/allochrt.pdf
[14] First Report and Order, Federal Communication Commission Std. FCC 02–48, Feb. 2002.
[15] H.B. Chang, K.C. Chen, N. Prasad, C.W. Su, 'Auction Based Spectrum Management for Cognitive Radio Networks', IEEE Vehicular Technology Conference, Spring, 2009.
[16] Q. Zhao, B.M. Sadler, 'A Survey of Dynamic Spectrum Access', *Signal Processing Magazine IEEE*, **24**(3), 2007, 79–89.
[17] S. Haykin, 'Cognitive Radio: Brain-empowered Wireless Communications', *IEEE Journal on Selected Areas in Communications*, **23**(2), 2005, 201–220.
[18] D. Niyato, E. Hossain, 'Competitive Pricing for Spectrum Sharing in Cognitive Radio Networks: Dynamic Game, Inefficiency of Nash Equilibrium, and Collusion', *IEEE Journal on Selected Areas in Communications*, **26**(1), 2008, 192–202.
[19] D. Niyato, E. Hossain, 'Competitive Spectrum Sharing in Cognitive Radio Networks: A Dynamic Game Approach', *IEEE Transactions on Wireless Communications*, **7**,July 2008, 2651–2660.
[20] Z. Ji, K.J.R. Liu, 'Multi-Stage Pricing Game for Collusion-Resistant Dynamic Spectrum Allocation', *IEEE Journal on Selected Areas in Communications*, **26**(1), 2008, 182–191.
[21] F. Wang, M. Krunz, S. Cui, 'Price-Based Spectrum Management in Cognitive Radio Networks', *IEEE Journal of Selected Topics in Signal Processing*, **2**(1), 2008, 74–87.
[22] S. Gjerstad, J. Dickhaut, 'Price Formation in Double Auctions', *Games and Economic Behavior*, **22**,1998, 1–29.
[23] V. Krishna, *Auction Theory*, Academic Press, 2002.
[24] N. Singh, X. Vives, 'Price and Quantity Competition in a Differentiated Duopoly', *RAND J. Economics*, **15**(4), 1984, 546–554.
[25] W. Hu, D. Willkomm, M. Abusubaih, J. Gross, G. Vlantis, M. Gerla, A. Wolisz, 'Cognitive Radio for Dynamic Spectrum Access – Dynamic Frequency Hopping Communities for Efficient IEEE 802.22 Operation', *IEEE Communications Magazine*, **45**(5), 2007, 80–87.
[26] M. Sherman, A.N. Mody, R. Martinez, C. Rodriguez, R. Reddy, 'IEE Standards Supporting Cognitive Radio and Networks, Dynamic Spectrum Access, and Coexistence', *IEEE Communications Magazine*, **46**(7), 2008, 72–79.
[27] I.F. Akyildiz, S. Mohanty, J. Xie, 'A Ubiquitous Mobile Communication Architecture for Next-Generation Heterogeneous Wireless Systems', *IEEE Communications Magazine*, **43**,2005, s29–s36.
[28] C.E. Perkins, 'Mobile Networking Through Mobile IP', *IEEE Internet Computing*, January/February 1998, 58–69.

国际视野　科技前沿

国际信息工程先进技术译丛

《认知无线电网络》
《高速数字系统的信号完整性和辐射发射》
《UMTS中的LTE：基于OFDMA和SC-FDMA的无线接入》
《生物医学工程学概论》（原书第2版）
《全面的功能验证：完整的工业流程》
《无线Mesh网络架构与协议》
《UMTS蜂窝系统的QoS与QoE管理》
《半导体制造与过程控制基础》
《WCDWA原理与开发设计》
《下一代移动系统：3G/B3G》
《IMS:IP多媒体概念和服务》（原书第2版）
《下一代无线系统与网络》
《深入浅出UMTS无线网络建模、规划与自动优化：理论与实践》
《HSDPA/HSUPA技术与系统设计——第三代移动通信系统宽带无线接入》
《无线传感器及元器件：网络、设计与应用》
《印制电路板——设计、制造、装配与测试》
《IPTV与网络视频：拓展广播电视的应用范围》
《多电压CMOS电路设计》
《微电子技术原理、设计与应用》
《蜂窝网络高级规划与优化2G/2.5G/3G/…向4G的演进》
《基于蜂窝系统的IMS——融合电信领域的VoIP演进》
《无线网络中的合作原理与应用》
《电生理学方法与仪器入门》
《移动电视：DVB-H、DMB、3G系统和富媒体应用》
《环境网络：支持下一代无线业务的多域协同网络》
《基于射频工程的UMTS空中接口设计与网络运行》
《未来UMTS的体系结构与业务平台：全IP的3G CDMA网络》
《UMTS-HSDPA系统的TCP性能》
《宽带无线通信中的空时编码》
《数字图像处理》（原书第4版）
《基于4G系统的移动服务技术》
《吉规模集成电路互连工艺及设计》
《高性能微处理器电路设计》

上架指导：工业技术/电信技术/无线通信



WILEY
www.wiley.com

地址：北京市百万庄大街22号
电话服务
社服务中心：(010)88361066
销 售 一 部：(010)68326294
销 售 二 部：(010)88379649
读者服务部：(010)68993821
邮政编码：100037
网络服务
门户网：http://www.cmpbook.com
教材网：http://www.cmpedu.com
封面无防伪标均为盗版

ISBN 978-7-111-32040-1
封面设计：马精明
定价：88.00元



